— Supplementary Figures —

Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases

Daniel Marbach[1,2,*], David Lamparter[1,2], Gerald Quon[3,4], Manolis Kellis[3,4], Zoltán Kutalik[1,2,5], Sven Bergmann[1,2]

[1] Department of Medical Genetics, University of Lausanne, Switzerland
[2] Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3] Broad Institute, MIT, Cambridge, MA, USA
[4] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA
[5] Institute of Social and Preventive Medicine, University Hospital of Lausanne, Switzerland

* Correspondence should be addressed to D.M. (daniel.marbach@unil.ch)

Website: http://regulatorycircuits.org

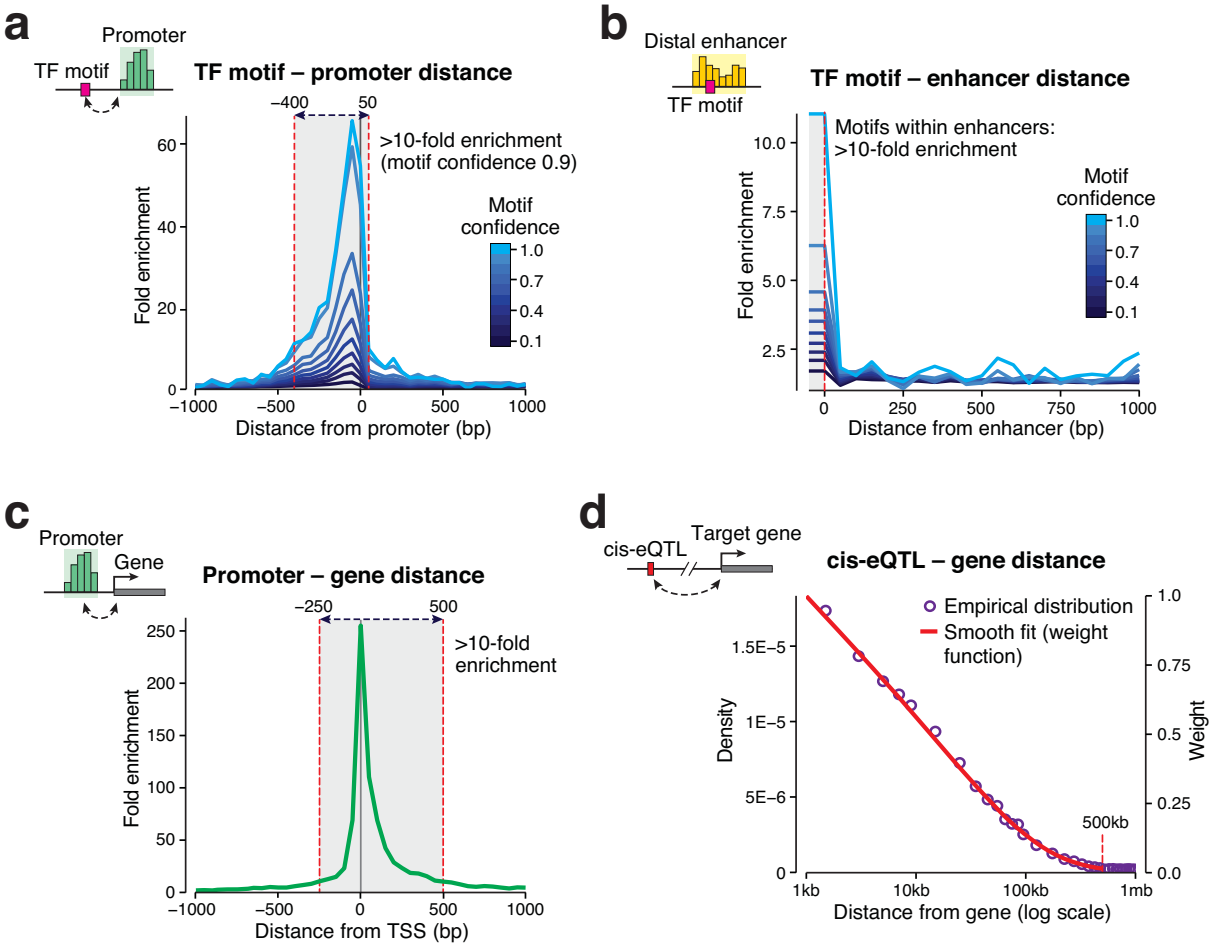
Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figures

Nature Methods: doi:10.1038/nmeth.3799

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 1



Supplementary Figure 1: Linking TFs, promoters, enhancers, and genes

(**a**) Enrichment of TF motif instances near CAGE-defined promoters. High-confidence motifs are >10-fold enriched in a window of -400–50bp around promoters (negative: upstream; positive: downstream).

(**b**) Enrichment of TF motif instances near distal, CAGE-defined enhancers. High-confidence motifs are >10-fold enriched inside (distance = 0bp) enhancers. In contrast to promoters, motif enrichment only occurs inside — not around — the CAGE-defined enhancers, indicating that the signature of bidirectional transcription used for their detection comprises the entire enhancer region.

(**c**) Enrichment of CAGE-defined promoters near the annotated transcription start site (TSS) of genes. Promoters are >10-fold enriched in a window of -250–500bp around the TSS of genes.

(**a-c**) Regions where >10-fold enrichment was found (grey areas) are used to link TF motifs to promoters, TF motifs to enhancers, and promoters to genes.

(**d**) Distance distribution of cis-eQTLs from their target genes (note: logarithmic scale on x-axis, orientation [up/downstream] is not distinguished). The smooth fit of the empirical distribution is used as a weighting function to probabilistically link enhancers to genes (second y-axis).

See also Methods for a more detailed discussion of these analyses.

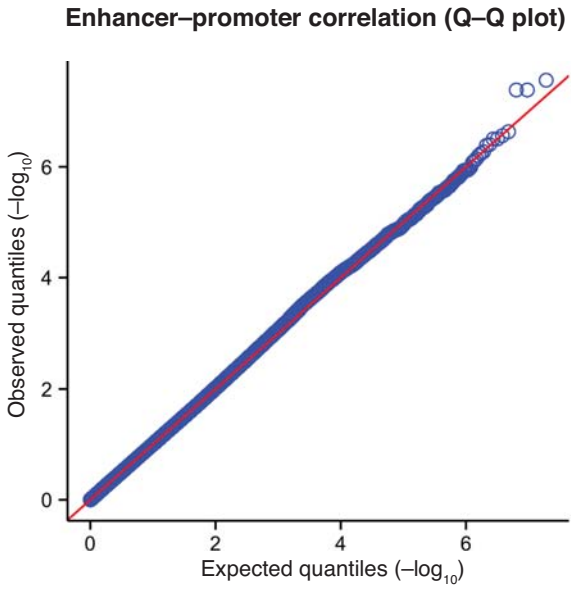
Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 2

Supplementary Figure 2: Enhancer–promoter activity correlations for chromosome 22
In order to investigate whether enhancers globally (i.e., across different cell types and tissues) associate with specific target promoters, we computed Pearson's correlation coefficient for all enhancer–promoter pairs on a given chromosome. The correlation matrix was visualized as a heatmap, where promoters and enhancers were ordered according to their position on the chromosome. Here, chromosome 22 is shown as an illustrative example, the same observations were made for other chromosomes. We found that most enhancers do not globally correlate more strongly with specific, nearby promoters (blue cells are scattered through the matrix). Instead, most enhancers positively correlate with many promoters throughout the chromosome, many of which are too far away to be likely regulatory targets of the enhancer. The reason is that many promoters are themselves co-expressed. Moreover, cell type-specific interactions are often missed when using correlation or correlation-like statistics[1] (this conclusion is also supported by a genome-wide analysis in 3). It is possible that more sophisticated approaches (e.g., based on an initial clustering of the input data,[2,3] statistics designed to capture cell type-specific interactions,[1] or machine learning approaches integrating additional data[4]) would reveal significant relationships. However, the performance of these different methods for the data at hand is not yet well understood and we thus opted for a parsimonious approach to link enhancers to potential target promoters (see Methods).

Nature Methods: doi:10.1038/nmeth.3799

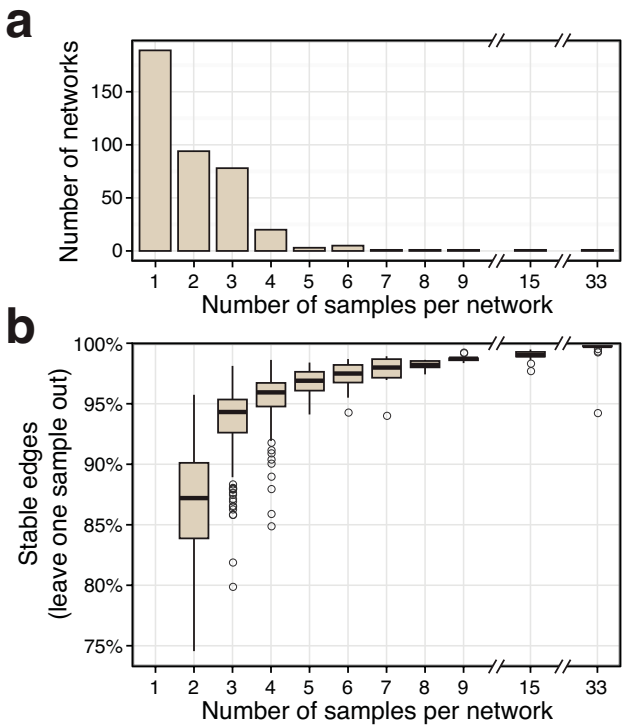
Supplementary Figure 3



Supplementary Figure 3: Q-Q plot for enhancer–promoter correlation coefficients
We sought to test whether enhancers show increased correlation with their target promoters. To this end, we computed Pearson's correlation coefficient for all enhancer–promoter pairs. We divided the correlation coefficients in two sets: one for enhancer–promoter pairs that are on the same chromosome (intra-chromosomal pairs), and one for pairs where the enhancer and promoter are on different chromosomes (inter-chromosomal pairs). Since the target promoters of an enhancer are on the same chromosome, the correlation coefficients corresponding to true enhancer–target promoter interactions are all in the intra-chromosomal set, while the inter-chromosomal correlation coefficients correspond to the expected null distribution for non-interacting enhancer–promoter pairs. Thus, if enhancers showed increased correlation with their target promoters, the first distribution would be shifted towards greater positive correlation. The Q-Q plot shows that this is not the case: the two distributions are identical. We conclude that enhancers do not show increased correlation with specific target promoters at a global level, i.e., across cell types and tissues (note that the FANTOM5 data consists of samples from different cell types and tissues — naturally, the situation would be different for a dataset consisting of many samples from the same cell type or tissue). See **Supplementary Fig. 2** for a more detailed discussion of these observations. (Note, for computational reasons the Q-Q plot only includes enhancer–promoter pairs for chromosomes 13–22; results are identical for the remaining chromosomes).

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 4



Supplementary Figure 4: Dependence of networks on individual samples

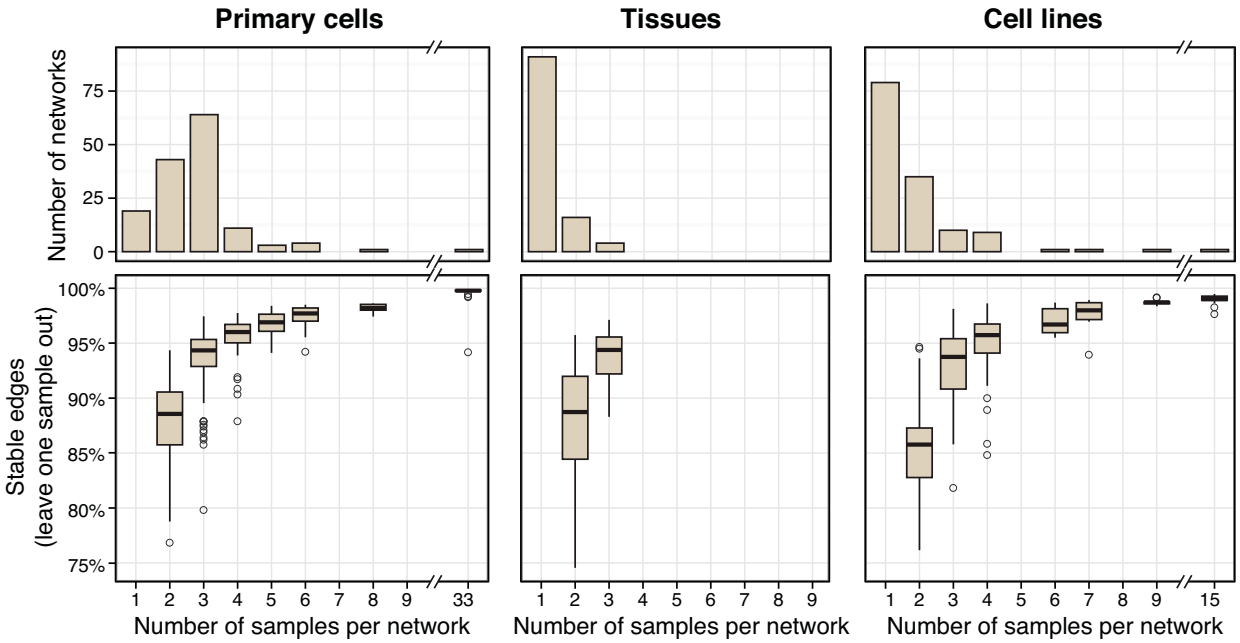
The compendium of 394 cell type and tissue-specific regulatory networks is based on 808 cell and tissue samples from FANTOM5: first a network was constructed for each of the 808 samples, and then networks of biological replicates and other closely related samples were merged (Supplementary Table 1, Methods).

(**a**) The histogram shows the number of samples merged for each of the 394 final networks. Most networks consist of few samples: 361 networks (92%) merge only one, two or three samples. 20 networks consist of 4 samples and only 13 networks have more than 4 samples.

(**b**) Dependence of networks on individual samples. For each of the networks consisting of more than one sample, we assessed the dependence on each of its samples. To this end, we removed one sample at a time and evaluated the percentage of edges that remained in the network (referred to as stable edges). Boxplots show the percentage of stable network edges when removing each of the samples (Supplementary Table 1 includes the result for each sample). These results confirm that the merged samples of a network are generally very similar and redundant to each other: e.g., for networks consisting of three samples, on average 94% of edges remain the same when leaving one of the samples out.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 5

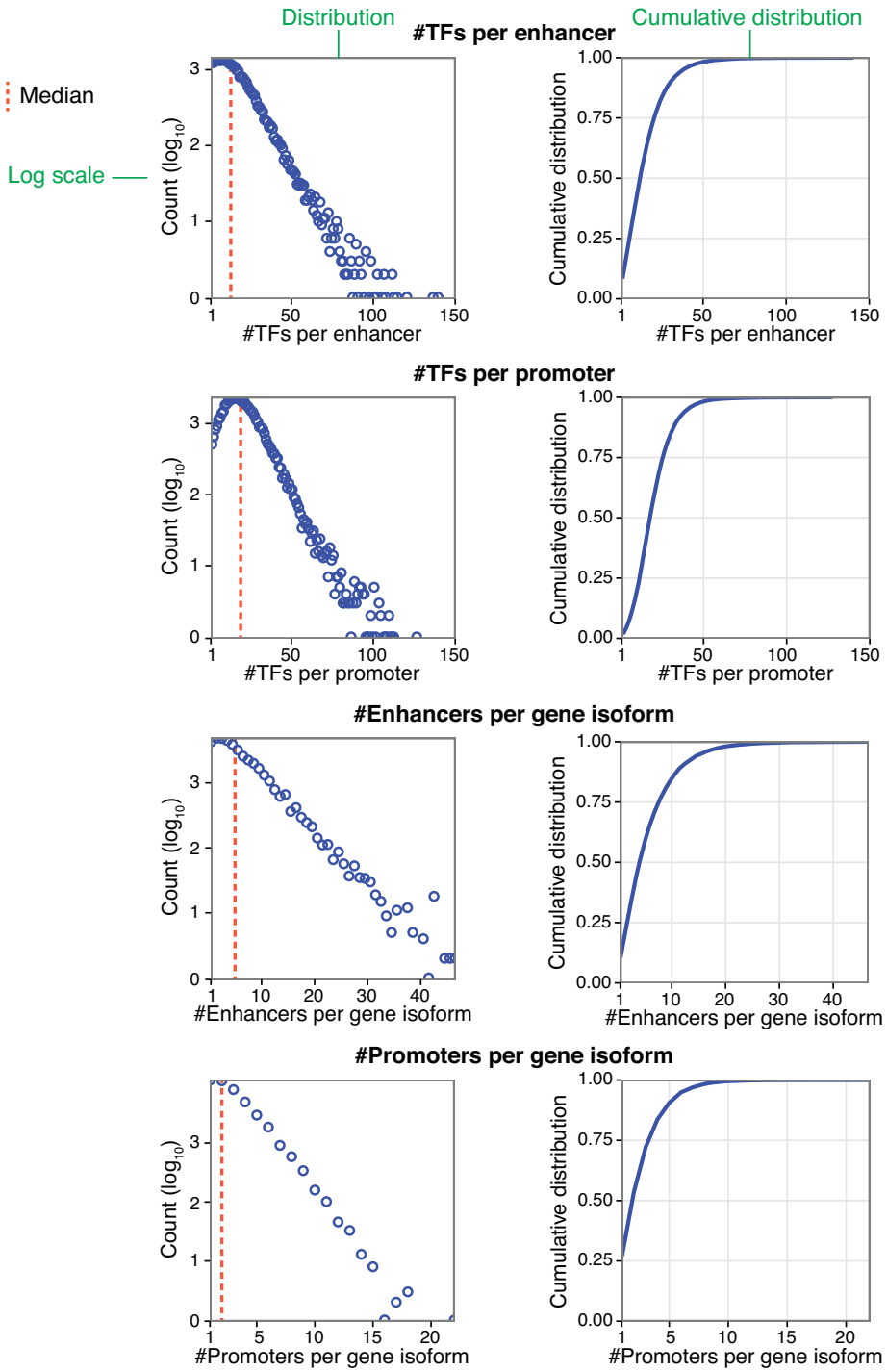


Supplementary Figure 5: Dependence of networks on individual samples for primary cells, tissues, and cell lines

Top row: histograms showing the number of samples merged per network; bottom row: dependence of networks on individual samples. For explanation, see legend of **Supplementary Fig. 4**. Here, the same results are shown broken down by primary cells, tissues, and cell lines. Stability of edges is slightly lower for cell lines than for primary cells and tissues. This is because merged primary cell or tissue samples are mostly biological replicates (same cell type or tissue from different donors), while merged cell lines are merely of the same cancer subtype (tumor samples from different patients are expected to have higher heterogeneity than healthy tissue samples).

Nature Methods: doi:10.1038/nmeth.3799

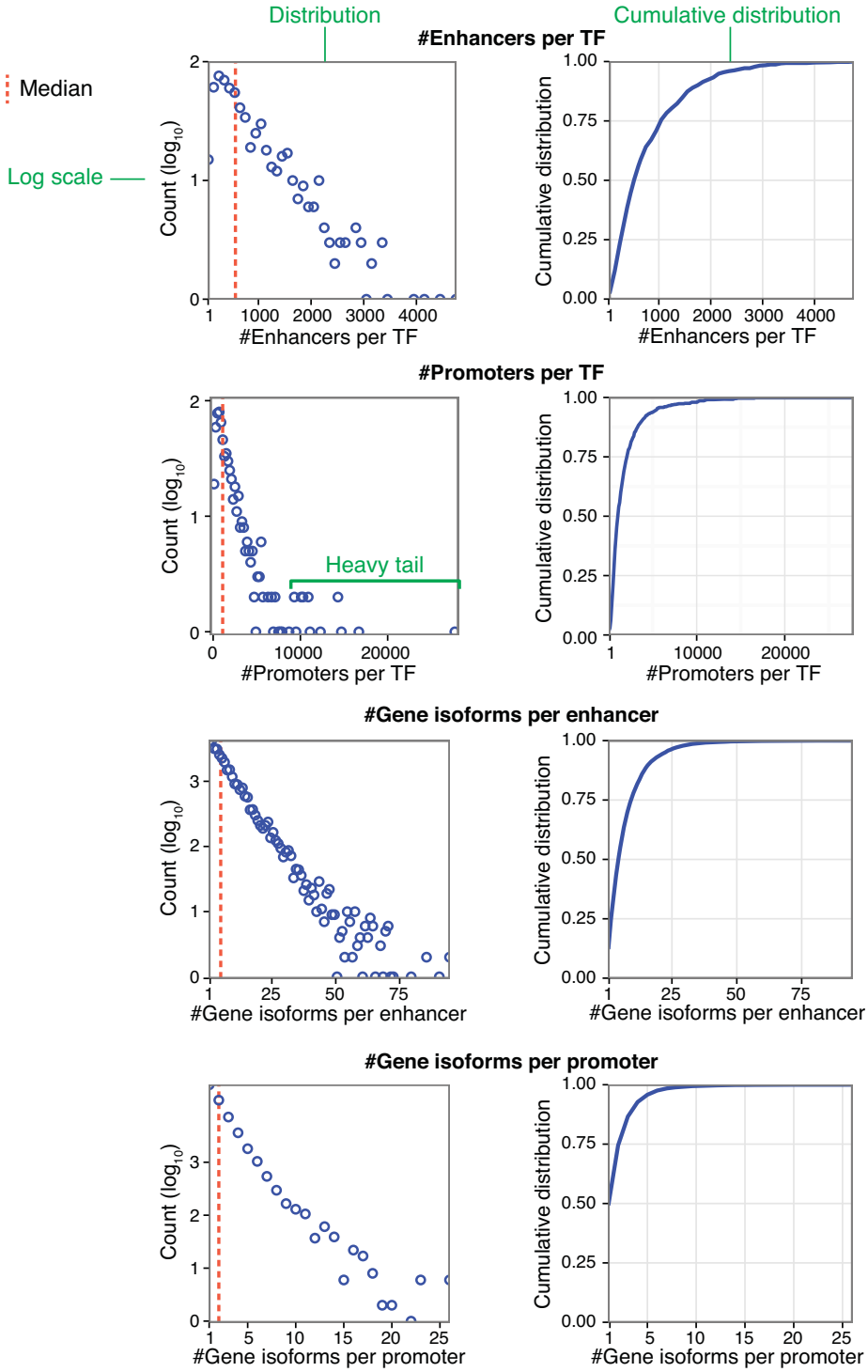
Supplementary Figure 6



Supplementary Figure 6: In-degree distributions of regulatory circuits at each layer
In-degree distributions (left side) show the number of nodes (y-axis) for each in-degree (number of incoming links, x-axis). Distributions are plotted with a logarithmic scale on the y-axis, revealing approximately exponential decay (linear relationship in the semi-log plots). Exponential in-degree distributions, which do not have extreme hubs as in scale-free networks, have been previously observed for gene networks.[5,6] Presumably, this is because it would be difficult to sensibly integrate a huge number of TF inputs at a single enhancer or promoter (a different phenomenon are hot regions where huge numbers of TFs bind non-specifically, but likely without direct regulatory effect[7]). Extreme hubs are also not expected for the other types of links (enhancer–gene and promoter–gene) because they are confined to a certain region on a chromosome, i.e., the number of potential interaction partners is physically limited.

Nature Methods: doi:10.1038/nmeth.3799

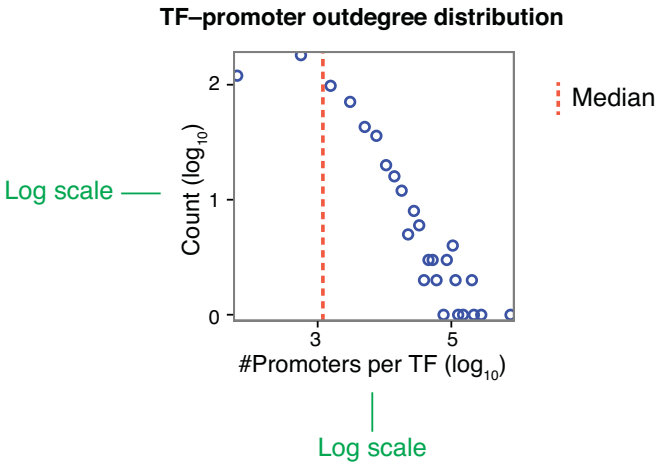
Supplementary Figure 7



Supplementary Figure 7: Out-degree distributions of regulatory circuits at each layer
Out-degree distributions (left side) show the number of nodes (y-axis) for each out-degree bin (number of outgoing links, counts were obtained using equally spaced bins on x-axis). Similar to the in-degree distributions (previous figure), the out-degree distributions show roughly exponential decay with one exception: the number of promoters per TF has a heavy tail corresponding to extreme hubs that are characteristic of scale-free networks (cf. **Supplementary Fig. 8**). In other words, regulatory networks are scale-free only at the level of promoters, but not at the level of enhancers, suggesting that the widely cited scale-free property of gene networks[6] is due to the promoter-centric view of previous studies. Concerning the enhancer–gene and promoter–gene links, we do not expect extreme hubs because of physical constraints, as discussed in **Supplementary Fig. 6**.

Nature Methods: doi:10.1038/nmeth.3799

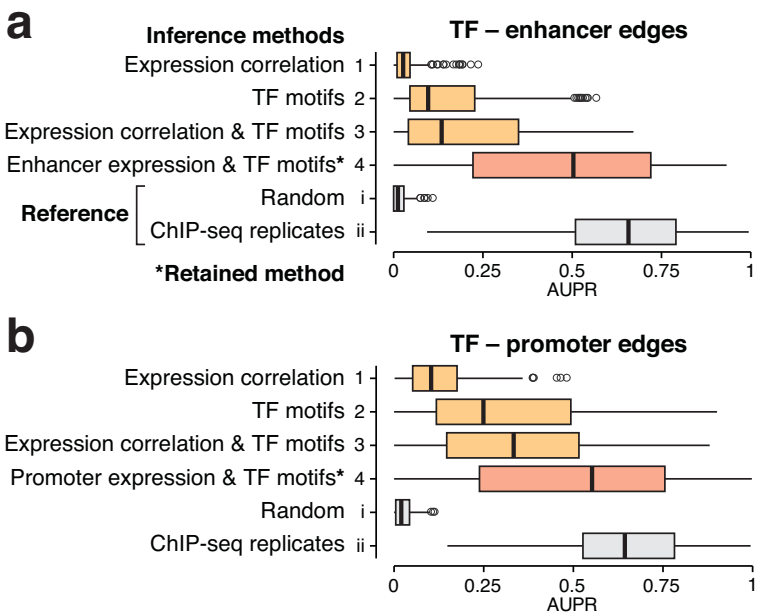
Supplementary Figure 8



Supplementary Figure 8: TF–promoter out-degree distribution
Same plot as shown in the second row of **Supplementary Fig. 7**, but using a log scale for both axes, confirming that the out-degree distribution of TFs with respect to promoters follows roughly a power-law in the right tail (linear relationship in the log-log plot).

Nature Methods: doi:10.1038/nmeth.3799

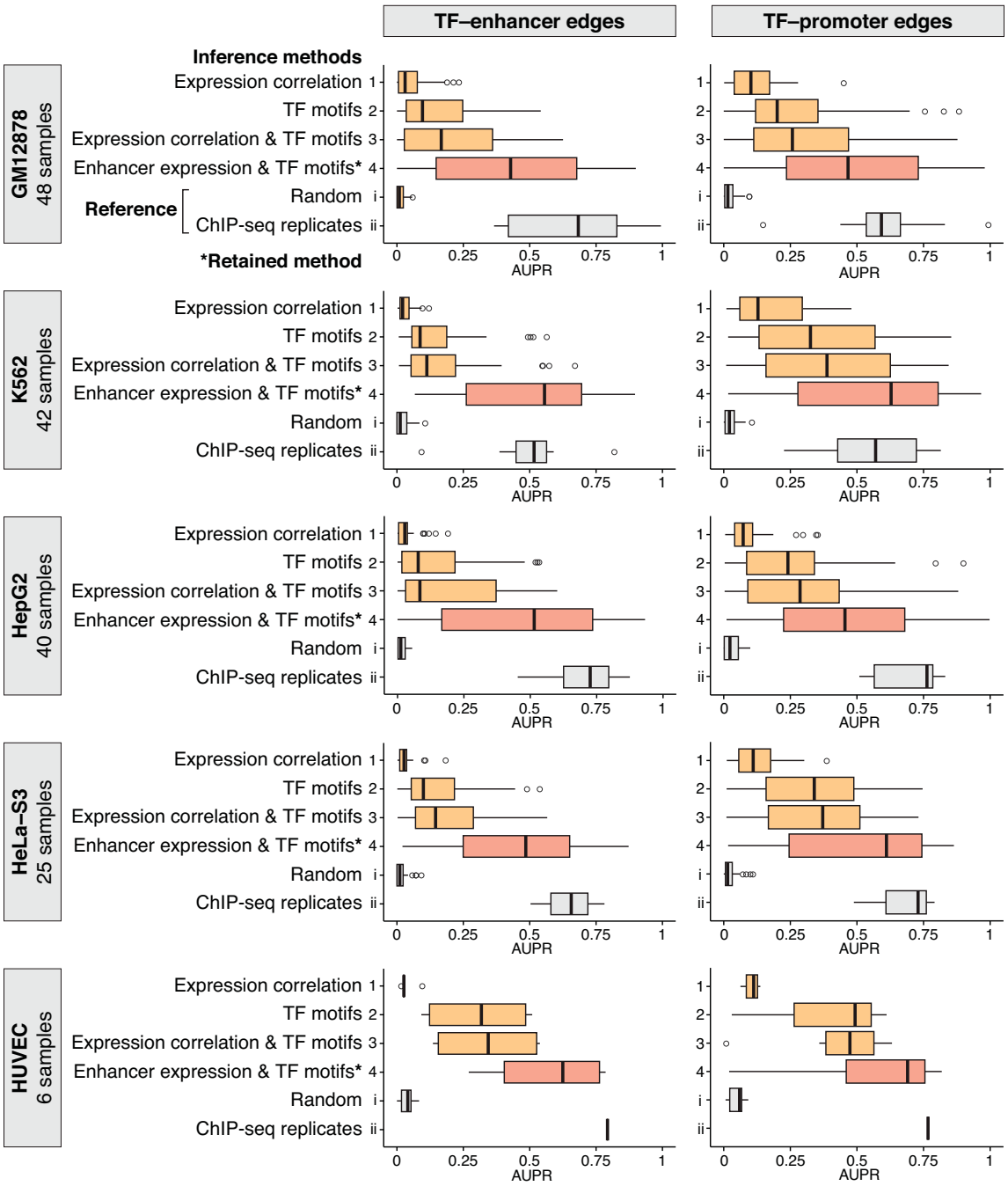
Supplementary Figure 9



Supplementary Figure 9: ChIP-seq validation of TF–enhancer and TF–promoter edges
Assessment of different approaches to infer edges between TFs and regulatory elements (enhancers and promoters). See legend of Fig. 2a for details. While Fig. 2a shows the overall performance across enhancers and promoters, here we show results individually for (**a**) enhancers and (**b**) promoters. For both the retained method to construct regulatory circuits (*) and the ChIP-seq replicates, there is no significant difference in performance between enhancers and promoters ($p > 0.05$, two-sided Wilcoxon rank-sum test).

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 10

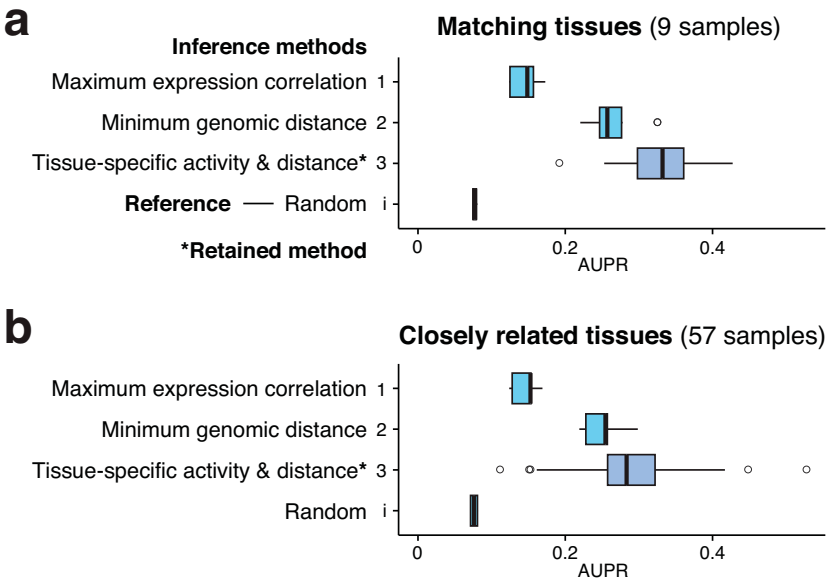


Supplementary Figure 10: ChIP-seq validation results for each of 5 cell lines

Assessment of different approaches to infer edges between TFs and regulatory elements (enhancers and promoters) for each of 5 ENCODE cell lines. See legend of Fig. 2a for details. While Fig. 2a shows the overall performance across enhancers, promoters and cell lines, here we show results individually for TF–enhancer and TF–promoter edges (columns) and the different cell lines (rows). For each cell line, the number of samples (ChIP-seq experiments) is indicated. Note that there are only 6 ChIP-seq experiemnts for HUVEC (human umbilical vein endothelial cells). The retained method to construct regulatory circuits (*) outperforms alternative approaches in each of the cell lines, but there is substantial variation in performance both for the retained method and ChIP-seq replicates across cell lines. Notably, the retained method performs as well as the ChIP-seq replicates in K562. There is also substantial variation in performance for different TFs in the same cell line, ranging from poor (AUPR comparable to random predictions) to excellent prediction accuracy (AUPR comparable to ChIP-seq replicates), which is expected due to varying quality of TF motifs and/or ChIP-seq antibodies. Comparing the performance for TF–enhancer vs. TF–promoter edges, we observe the same trends as in **Supplementary Fig. 9**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 11



Supplementary Figure 11: eQTL validation of enhancer–gene edges

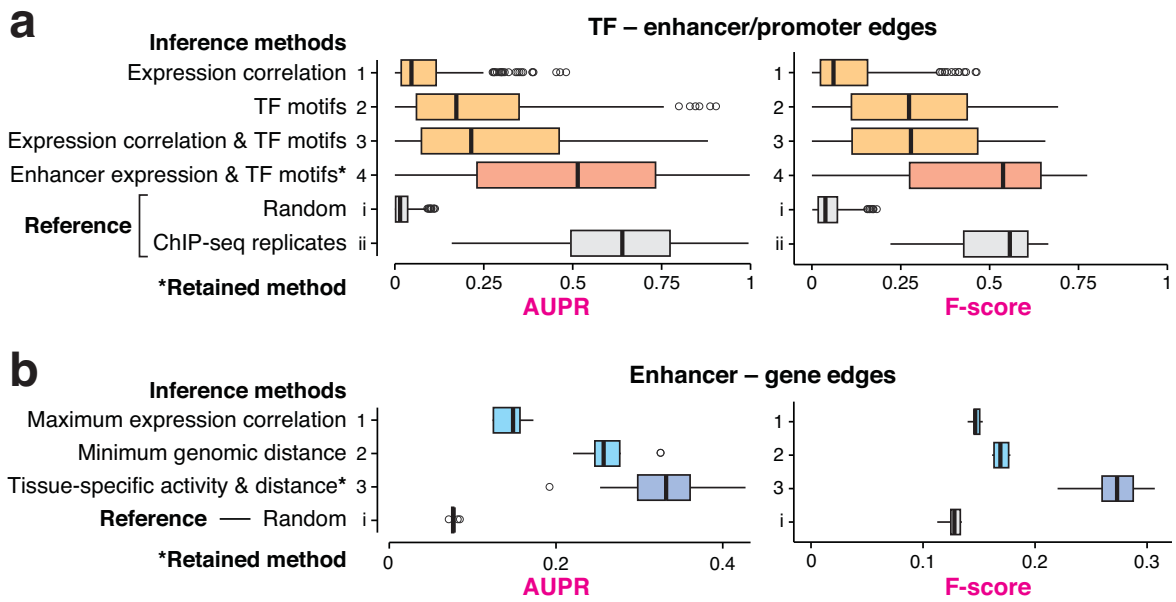
Assessment of different approaches to link enhancers to target genes using eQTL data from GTEx. See legend of Fig. 2b for details. We mapped the 13 tissues from GTEx to corresponding tissue samples from FANTOM (Supplementary Table 3). We distinguished between exact matches and closely related tissues. For example, one of the GTEx tissues is the aorta. There is a matching tissue sample in FANTOM, but also several aortic cell types including aortic endothelial cells, fibroblasts and smooth muscle cells, which we label as *closely related tissues*.

(**a**) Results for 9 matching GTEx–FANTOM tissue pairs (this panel is identical to Fig. 2b).

(**b**) Results for 57 closely related GTEx–FANTOM tissue pairs. The results for matching tissues are confirmed in the broader set of closely related tissues. As expected, for our tissue-specific circuits (the retained method) performance is slightly better for matching tissues than for closely related tissues. For the other methods (random, maximum correlation, and minimum distance), there is no difference because these methods do not infer links in a tissue-specific manner.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 12

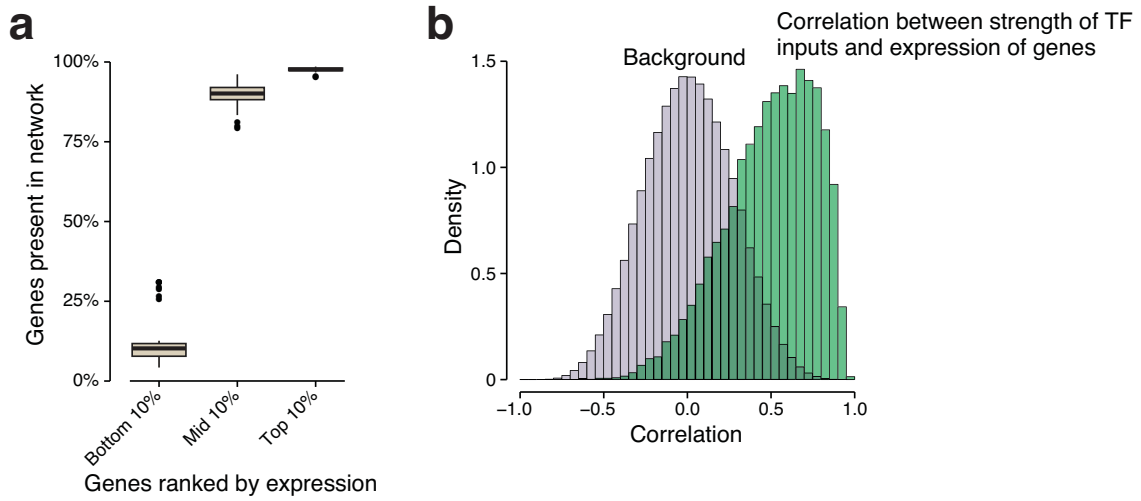


Supplementary Figure 12: ChIP-seq and eQTL validation: AUPR vs. F-score

Comparison of validation results using two alternative metrics to assess the accuracy of edge predictions: the area under the precision-recall curve (AUPR, left) and the F-score (right). The plots showing the AUPR are identical to Fig. 2 of the main text, they are reproduced here for comparison with the F-score. Refer to the legend of Fig. 2 for explanation. The choice of the AUPR to report validation results is motivated in Methods, but similar results were obtained with other performance metrics. Here we show in addition the F-score of edge predictions: the F-score is a popular metric used in the network inference literature, it is defined as the harmonic mean of precision and recall at a given cutoff. The plots show the F-score for the top 50% of the predicted edges, similar results were obtained at different cutoffs. We conclude that our results are robust and not specific to the AUPR as performance metric.

Nature Methods: doi:10.1038/nmeth.3799

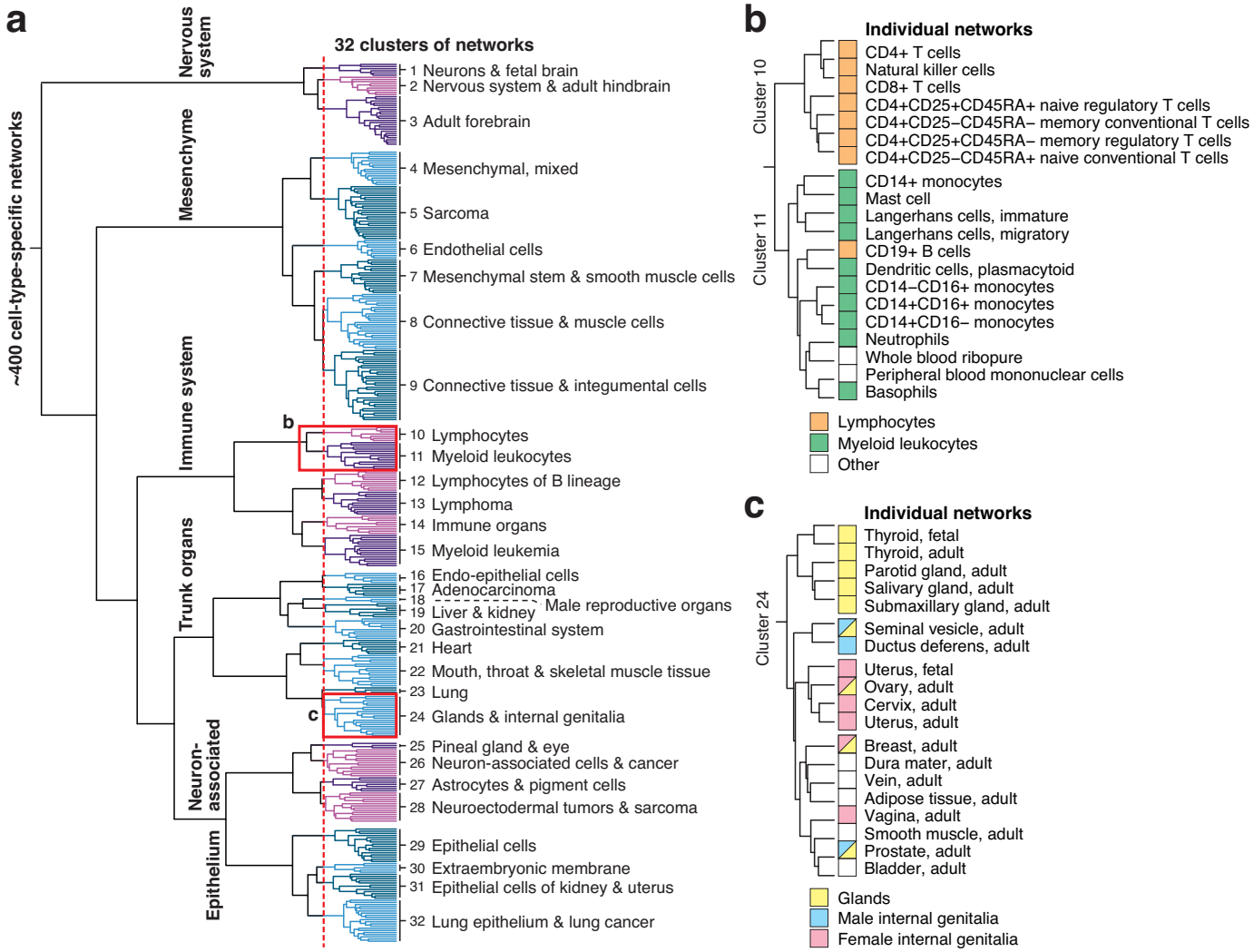
Supplementary Figure 13



Supplementary Figure 13: Correlation between regulatory edges and target gene expression
Gene expression data (RNA-seq) was obtained from the Roadmap Epigenomics project for 40 tissues that are also present in our network compendium (Supplementary Table 4, Methods).
(**a**) Overlap of weakly, moderately, and highly expressed genes with regulatory networks across the 40 tissues. For each tissue, genes were ranked by their expression level (RPKM). Boxplots show the percentage of genes that are part of the regulatory network for the bottom 10%, mid 10%, and top 10% most highly expressed genes. On average, 98% of the highly and 90% of the moderately expressed genes are present in our networks, i.e., they have regulatory inputs (incoming edges) in the given tissue. As expected, most of the weakly expressed genes have no active regulatory inputs in the given tissue, i.e., they are not present in the corresponding network.
(**b**) Correlation between regulatory edge strength and target gene expression. For each gene, the correlation between its TF input (sum of weights of incoming edges) and expression was evaluated across tissues. The plot shows the distribution of correlation coefficients for all genes in green. The background/expected distribution is shown in grey (correlation coefficients for TF inputs of gene i and expression levels of gene j, where $i \neq j$). There is strong positive correlation between the strength of TF inputs and the expression of the same gene. The plot shows results using Spearman's correlation coefficient (median = 0.53), similar results were obtained using Pearson correlation (median = 0.49). Correlation is stronger when considering only Roadmap Epigenomic tissues that were considered to be a good match for a network in our compendium (results shown here) than when including also the secondary matches / related tissues given in Supplementary Table 4 (median Spearman's correlation = 0.42).

Nature Methods: doi:10.1038/nmeth.3799

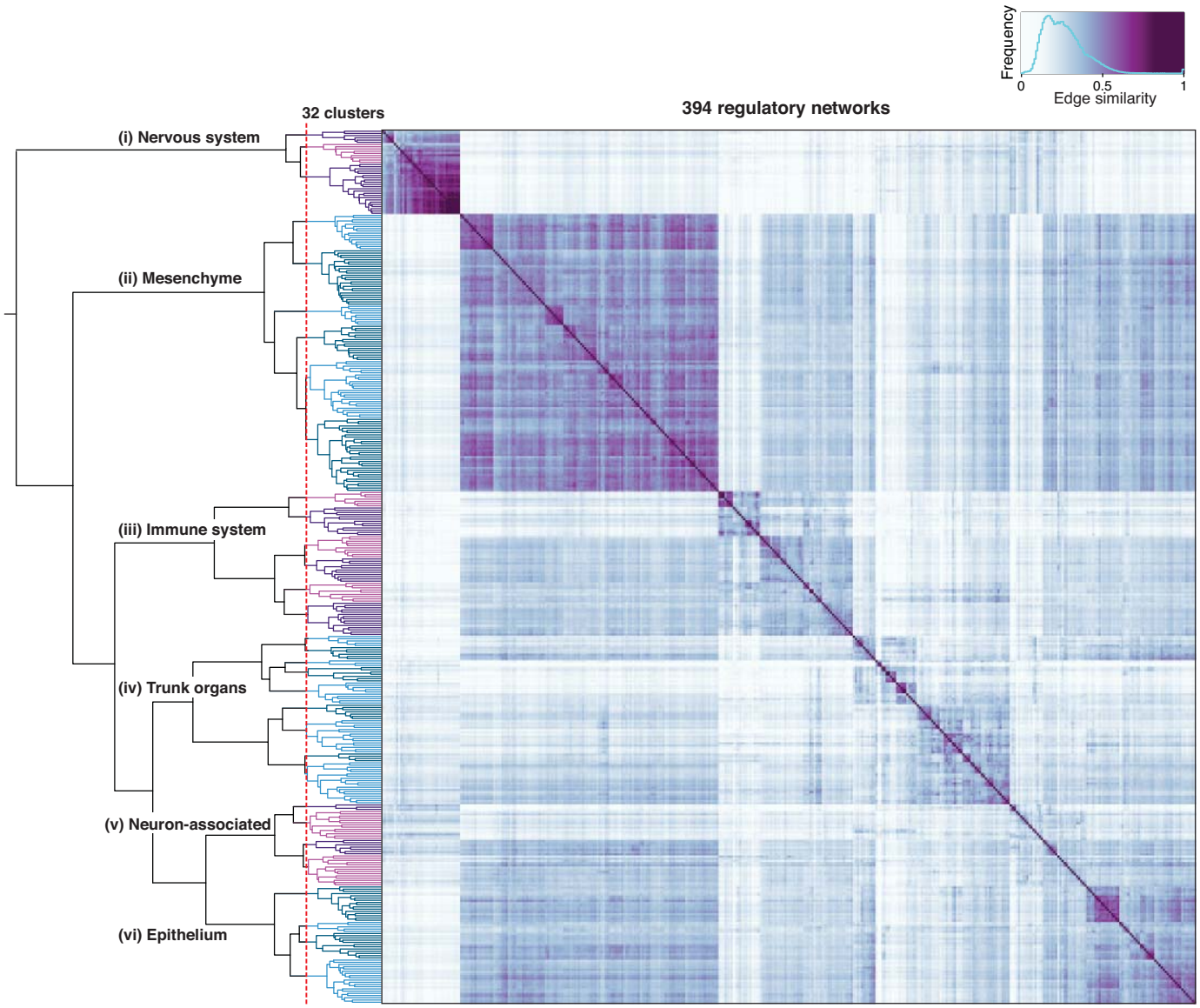
Supplementary Figure 14



Supplementary Figure 14: Hierarchical clustering of regulatory networks across cell types and tissues
(**a**) Clustering of the 394 cell type and tissue-specific regulatory networks based on edge similarity (see also **Supplementary Fig. 15**). Clusters are highly consistent with developmental and functional relationships between cell types, tissues and organs. The cutoff indicated by the red dashed line leads to 32 coherent clusters, annotated to the right based on enriched terms from cell, anatomical and disease ontologies. For each of these clusters, a high-level network was derived by merging the corresponding individual networks (Methods).
(**b**-**c**) Detailed view of representative clusters of networks (red boxes in Panel a; remaining clusters are shown in Supplementary Figs. **16**–**23**).
(**b**) Cluster 10 consists of regulatory networks from lymphocytes including T cells and natural killer cells, while Cluster 11 mostly contains networks from myeloid leukocytes. Cf. **Supplementary Fig. 19**.
(**c**) Cluster 24 joins regulatory networks from glands and internal genitalia (seminal vesicle, ovary, breast, and prostate belonging to both categories). Functionally related networks are consistently grouped also at the sub-cluster level (e.g., endocrine glands, salivary glands, and male and female genitalia, respectively). Cf. Supplementary Figs. **20**–**21**

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 15

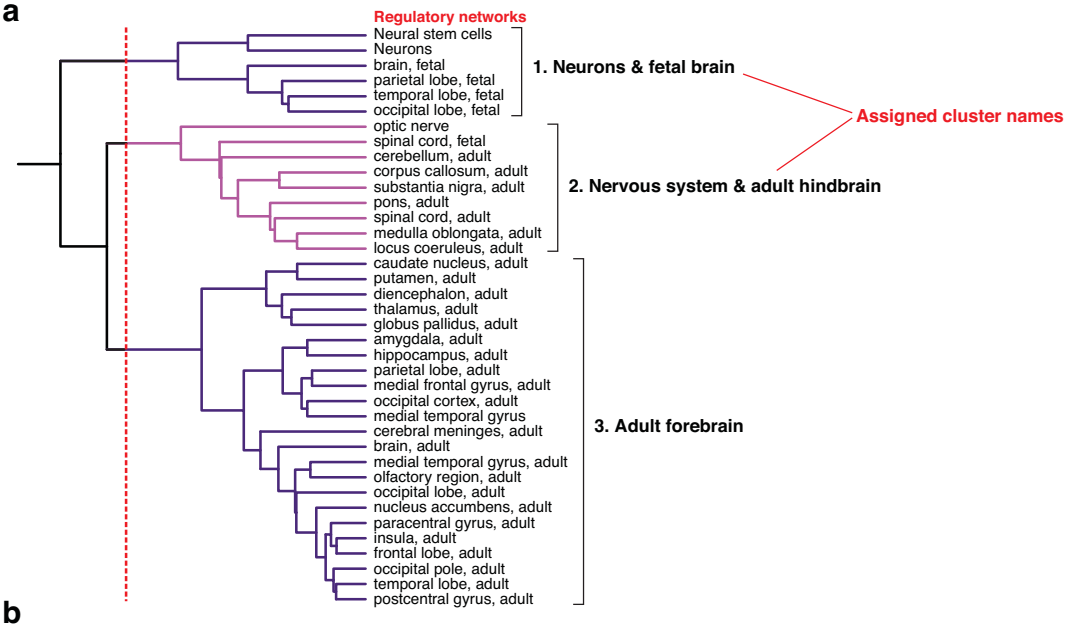


Supplementary Figure 15: Pairwise similarity of cell type and tissue-specific regulatory networks Hierarchical clustering of regulatory networks was performed based on edge similarity (Methods). The cutoff indicated by the red, dashed line leads to 32 coherent clusters, which are annotated in **Supplementary Fig. 14** and shown in detail below (Supplementary Figs. **16**–**23**). The six high-level groups (i–vi) separate regulatory networks of major types of tissues and anatomical systems, which also share distinct developmental origins (**Supplementary Fig. 24**). Regulatory networks of the nervous system form a particularly tight group and are the most dissimilar to networks of other groups.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 16

a



b

1. Neurons, fetal brain

Q-value	ID	Term
3.57E-02	UBERON:0000922	embryo
9.66E-02	UBERON:0011216	organ system subdivision

2. Nervous system, hindbrain

Q-value	ID	Term
1.95E-06	UBERON:0001016	nervous system
3.14E-06	UBERON:0004732	segmental subdivision of nervous system
3.59E-05	UBERON:0001017	central nervous system
5.29E-05	UBERON:0010314	structure with developmental contribution from neural crest
1.07E-04	UBERON:0011216	organ system subdivision
1.11E-04	UBERON:0002028	hindbrain
1.11E-04	UBERON:0004733	segmental subdivision of hindbrain
2.70E-04	UBERON:0004121	ectoderm-derived structure
1.03E-03	UBERON:0002298	brainstem
1.86E-03	UBERON:0000073	regional part of nervous system
1.86E-03	UBERON:0002616	regional part of brain
3.57E-03	UBERON:0001895	metencephalon
3.57E-03	UBERON:0002680	regional part of metencephalon
6.00E-03	UBERON:0000955	brain
1.81E-02	UBERON:0000477	anatomical cluster
3.13E-02	UBERON:0000122	neuron projection bundle
3.13E-02	UBERON:0000988	pons
3.13E-02	UBERON:0001137	dorsum
3.13E-02	UBERON:0002240	spinal cord
3.13E-02	UBERON:0005162	multi cell component structure
3.13E-02	UBERON:0005174	back organ
6.45E-02	UBERON:0000063	organ segment
7.64E-02	UBERON:0000481	multi-tissue structure
2.96E-03	UBERON:0007023	adult organism

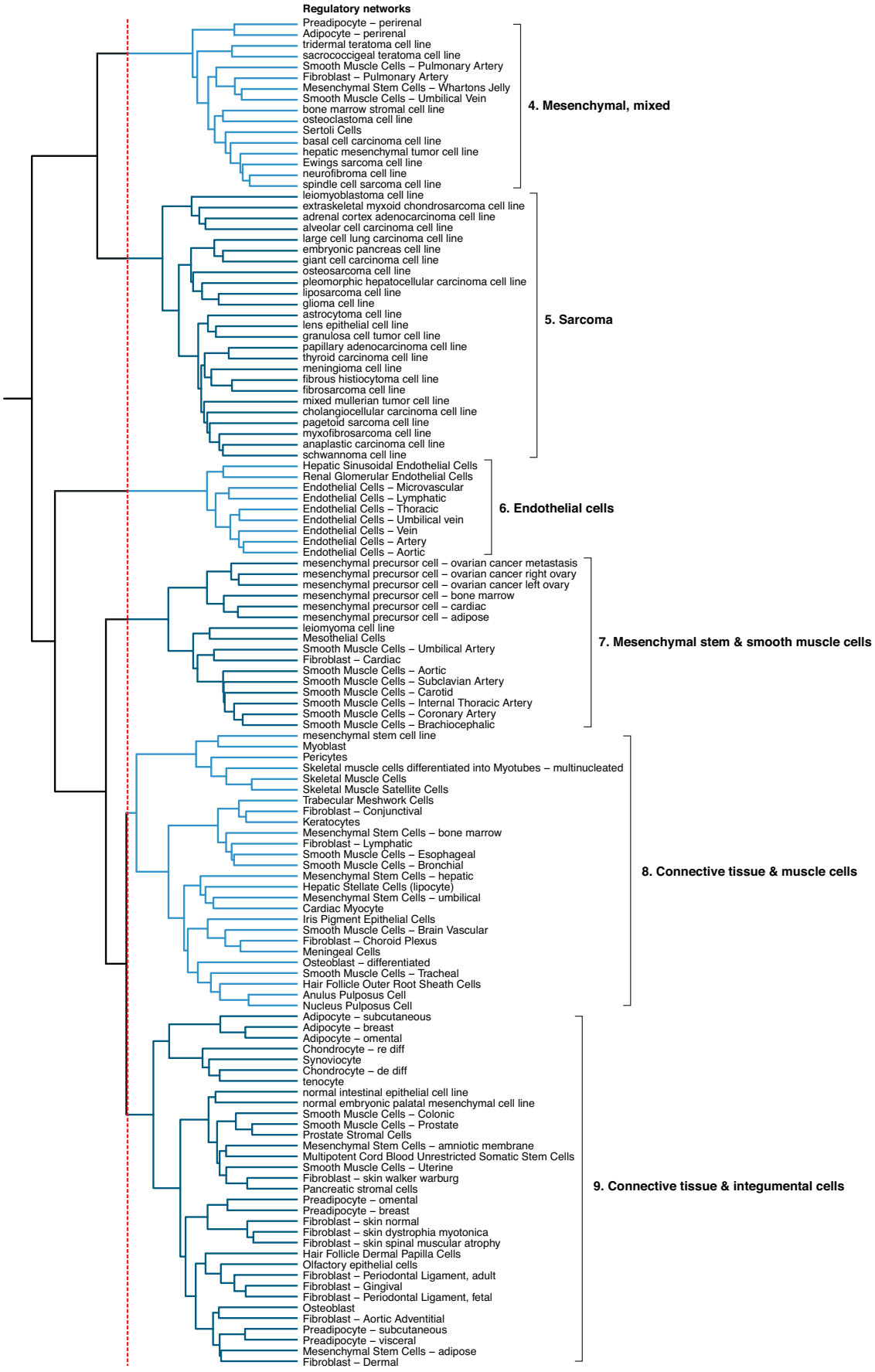
3. Adult forebrain

Q-value	ID	Term
1.68E-20	UBERON:0001890	forebrain
1.68E-20	UBERON:0002780	regional part of forebrain
1.87E-19	UBERON:0001869	cerebral hemisphere
9.13E-19	UBERON:0001017	central nervous system
9.90E-19	UBERON:0000073	regional part of nervous system
9.90E-19	UBERON:0002616	regional part of brain
1.44E-18	UBERON:0001893	telencephalon
2.79E-18	UBERON:0000955	brain
7.92E-18	UBERON:0001016	nervous system
1.24E-17	UBERON:0002791	regional part of telencephalon
3.90E-17	UBERON:0002020	gray matter of neuraxis
3.90E-17	UBERON:0003528	brain grey matter
1.87E-14	UBERON:0011216	organ system subdivision
6.79E-13	UBERON:0002619	regional part of cerebral cortex
2.50E-12	UBERON:0000203	pallium
2.50E-12	UBERON:0000956	cerebral cortex
1.04E-11	UBERON:0010314	structure with developmental contribution from neural crest
1.52E-11	UBERON:0001950	neocortex
2.32E-11	UBERON:0004121	ectoderm-derived structure
5.60E-11	UBERON:0000481	multi-tissue structure
1.08E-07	UBERON:0000477	anatomical cluster
5.08E-05	UBERON:0000064	organ part
8.76E-05	UBERON:0000200	gyrus
3.58E-04	UBERON:0000454	cerebral subcortex
3.58E-04	UBERON:0002420	basal ganglion
3.58E-04	UBERON:0007245	nuclear complex of neuraxis
3.58E-04	UBERON:0010009	aggregate regional part of brain
3.58E-04	UBERON:0010011	collection of basal ganglia
4.49E-03	UBERON:0001871	temporal lobe
4.49E-03	UBERON:0009663	telencephalic nucleus
1.11E-02	UBERON:0000125	neural nucleus
1.11E-02	UBERON:0002308	nucleus of brain
1.48E-02	UBERON:0000204	ventral part of telencephalon
1.48E-02	UBERON:0000349	limbic system
1.48E-02	UBERON:0000369	corpus striatum
1.48E-02	UBERON:0002435	striatum
4.17E-02	UBERON:0002021	occipital lobe
3.06E-14	UBERON:0007023	adult organism

Supplementary Figure 16: Dendrogram and annotation of nervous system clusters
(**a**) Dendrogram showing networks and cluster names for (i) Nervous system (cf. **Supplementary Fig. 15**).
(**b**) Enriched annotations from cell, anatomical and disease ontologies for each cluster of networks (p-values were corrected using Benjamini-Hochberg procedure; Methods). Horizontal lines separate groups of related annotations. We named clusters by inspecting both the individual networks (**a**) and the enriched annotations (**b**, selected terms are highlighted in yellow) with the aim to define succinct names that fit the majority of corresponding cells and tissues.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 17



Supplementary Figure 17: Dendrogram of mesenchyme clusters

Dendrogram showing networks and cluster names for (ii) Mesenchyme (cf. **Supplementary Fig. 15**). Annotations are shown in **Supplementary Fig. 18**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 18

4. Mesenchymal, mixed

Q-value	ID	Term
8.15E-02	UBERON:0000078	mixed ectoderm/mesoderm/endoderm-derived structure
8.15E-02	UBERON:0002012	pulmonary artery
8.15E-02	UBERON:0008886	pulmonary vascular system
8.15E-02	UBERON:0005406	perirenal fat
4.40E-02	DOID:3307	teratoma

5. Sarcoma

Q-value	ID	Term
5.03E-02	UBERON:0000949	endocrine system
7.42E-02	UBERON:0002368	endocrine gland
6.91E-07	DOID:4	disease
2.70E-06	DOID:14566	disease of cellular proliferation
2.05E-05	DOID:162	cancer
1.36E-03	DOID:0050687	cell type cancer
5.06E-03	DOID:7	disease of anatomical entity
2.82E-02	DOID:1115	sarcoma
2.99E-03	DOID:0060100	musculoskeletal system cancer
2.99E-03	DOID:201	connective tissue cancer

6. Endothelial cells

Q-value	ID	Term
8.56E-09	UBERON:0001986	endothelium
8.56E-09	UBERON:0004638	blood vessel endothelium
8.56E-09	UBERON:0004852	cardiovascular system endothelium
1.40E-07	UBERON:0000055	vessel
5.90E-07	UBERON:0002049	vasculature
5.90E-07	UBERON:0007798	vascular system
7.99E-07	UBERON:0003914	epithelial tube
1.34E-06	UBERON:0000487	simple squamous epithelium
3.14E-06	UBERON:0001981	blood vessel
3.14E-06	UBERON:0004537	blood vasculature
4.70E-06	UBERON:0006914	squamous epithelium
5.70E-06	UBERON:0004111	anatomical conduit
1.15E-05	UBERON:0000025	tube
1.41E-05	UBERON:0004535	cardiovascular system
1.61E-05	UBERON:0001009	circulatory system
2.98E-05	UBERON:0000490	unilaminar epithelium
2.50E-04	UBERON:0000483	epithelium
2.77E-04	UBERON:0000119	cell layer
1.03E-03	UBERON:0001917	endothelium of artery
1.03E-03	UBERON:0003915	endothelial tube
1.03E-03	UBERON:0004700	arterial system endothelium
1.81E-02	UBERON:0000477	anatomical cluster
1.97E-02	UBERON:0004120	mesoderm-derived structure
6.12E-13	CL:0000115	endothelial cell
3.38E-11	CL:0000213	lining cell
3.38E-11	CL:0000215	barrier cell
3.38E-11	CL:0002078	meso-epithelial cell
8.91E-11	CL:0002139	endothelial cell of vascular tree
8.06E-09	CL:0000071	blood vessel endothelial cell
2.46E-04	CL:0000076	squamous epithelial cell
5.12E-04	CL:0000066	epithelial cell
7.81E-04	CL:1000413	endothelial cell of artery
2.76E-02	CL:0002543	vein endothelial cell
2.76E-02	CL:0002544	aortic endothelial cell

7. Mesenchymal stem & smooth muscle cells

Q-value	ID	Term
3.47E-03	UBERON:0003914	epithelial tube
2.52E-02	UBERON:0000025	tube
2.95E-05	UBERON:0001637	artery
2.95E-05	UBERON:0003509	arterial blood vessel
2.95E-05	UBERON:0004572	arterial system
6.36E-05	UBERON:0004571	systemic arterial system
6.36E-05	UBERON:0004573	systemic artery
3.58E-04	UBERON:0001981	blood vessel
3.58E-04	UBERON:0004537	blood vasculature
4.52E-04	UBERON:0004535	cardiovascular system
5.45E-04	UBERON:0001009	circulatory system
9.32E-04	UBERON:0000055	vessel
1.41E-03	UBERON:0004120	mesoderm-derived structure
2.75E-03	UBERON:0002049	vasculature
2.75E-03	UBERON:0007798	vascular system
8.15E-02	UBERON:0001533	subclavian artery
2.47E-03	CL:0000134	mesenchymal cell
3.39E-03	CL:0000048	multi fate stem cell
4.57E-03	CL:0000723	somatic stem cell
6.05E-03	CL:0000034	stem cell
8.60E-07	CL:0000359	vascular associated smooth muscle cell
5.53E-05	CL:0000192	smooth muscle cell
2.75E-04	CL:0000187	muscle cell
3.82E-04	CL:0000211	electrically active cell
3.82E-04	CL:0000393	electrically responsive cell
5.12E-04	CL:0000183	contractile cell
7.29E-02	CL:0002595	smooth muscle cell of the subclavian artery
1.44E-02	CL:0002494	cardiocyte
8.63E-03	DOID:2394	ovarian cancer

8. Connective tissue & muscle cells

Q-value	ID	Term
5.16E-02	UBERON:0001134	skeletal muscle tissue
5.16E-02	UBERON:0002036	striated muscle tissue
8.98E-02	UBERON:0002204	musculoskeletal system
6.29E-03	UBERON:0002384	connective tissue
7.08E-03	UBERON:0006876	vasculature of organ
2.06E-04	CL:0000183	contractile cell
8.11E-04	CL:0000187	muscle cell
1.14E-03	CL:0000211	electrically active cell
1.14E-03	CL:0000393	electrically responsive cell
1.62E-02	CL:0000188	skeletal muscle cell
1.28E-03	CL:0002320	connective tissue cell

9. Connective tissue & integumental cells

Q-value	ID	Term
2.80E-03	UBERON:0002199	integument
2.80E-03	UBERON:0002416	integumental system
6.29E-03	UBERON:0003102	surface structure
5.58E-02	UBERON:0002097	skin of body
2.32E-14	UBERON:0002384	connective tissue
1.86E-03	UBERON:0004923	organ component layer
3.05E-15	CL:0002320	connective tissue cell
8.51E-07	CL:0000057	fibroblast
1.39E-03	CL:0002620	skin fibroblast
1.32E-02	CL:0002334	preadipocyte
1.58E-02	CL:0000499	stromal cell
8.95E-02	CL:0000136	fat cell

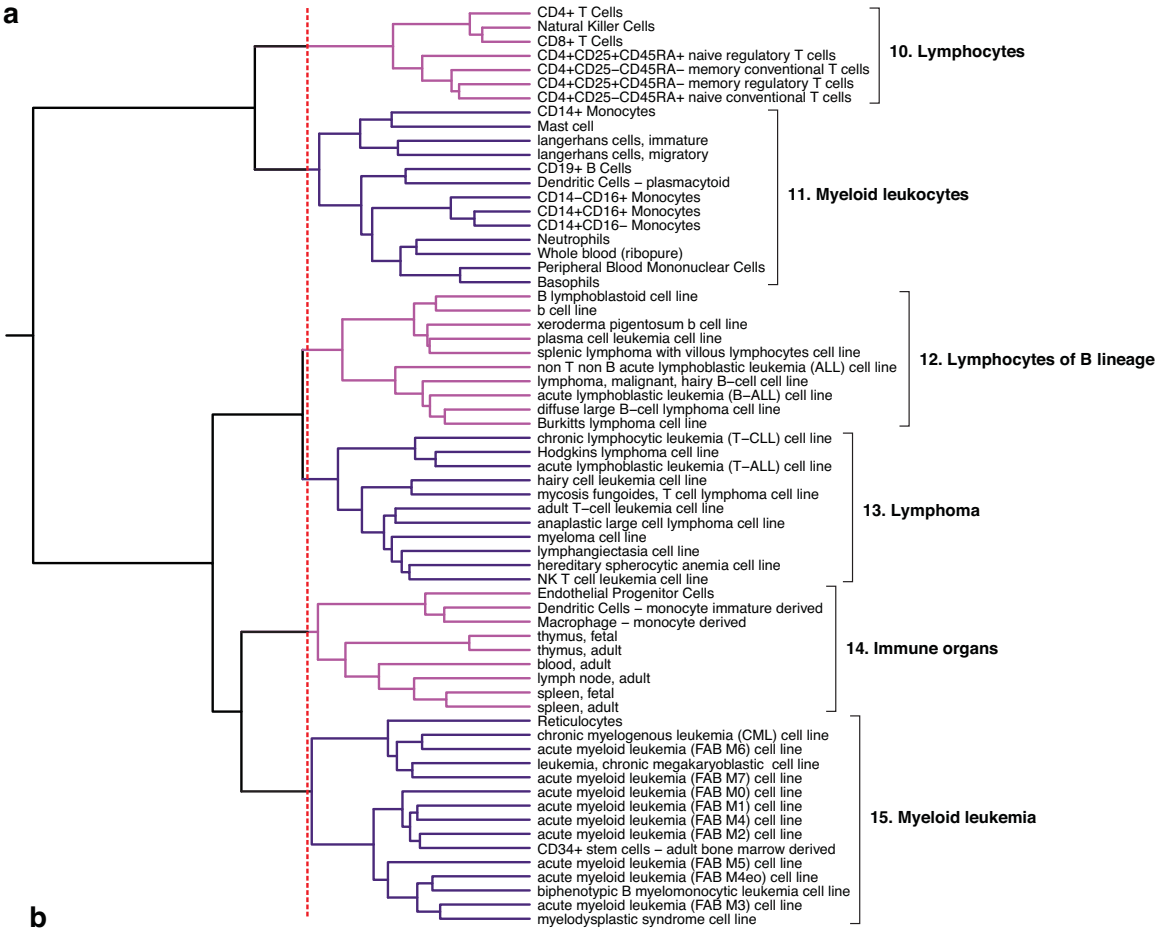
Supplementary Figure 18: Annotation of mesenchyme clusters

Enriched annotations for each cluster shown in **Supplementary Fig. 17**. Explanation in legend of **Supplementary Fig. 16**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 19

a



b

10. Lymphocytes

Q-value	ID	Term
1.14E-03	CL:0000789	alpha-beta T cell
1.14E-03	CL:0000791	mature alpha-beta T cell
1.14E-03	CL:0002419	mature T cell
1.50E-02	CL:0000542	lymphocyte
1.50E-02	CL:0002242	nucleate cell
1.69E-02	CL:0000624	CD4-positive, alpha-beta T cell
2.39E-02	CL:0000084	T cell
4.15E-02	CL:0002087	nongranular leukocyte

11. Myeloid leukocytes

Q-value	ID	Term
1.46E-09	CL:0000738	leukocyte
7.89E-08	CL:0000988	hematopoietic cell
1.74E-07	CL:0000219	motile cell
8.60E-07	CL:0000766	myeloid leukocyte
2.14E-04	CL:0000763	myeloid cell
9.46E-04	CL:0000576	monocyte
7.83E-03	CL:0000451	dendritic cell
7.83E-03	CL:0000990	conventional dendritic cell
7.86E-03	CL:0002087	nongranular leukocyte
5.09E-02	CL:0000094	granulocyte
5.09E-02	CL:0000453	Langerhans cell
5.09E-02	CL:0000860	classical monocyte
5.09E-02	CL:0002057	CD14-positive, CD16-negative classical monocyte
5.09E-02	CL:0002393	intermediate monocyte
5.09E-02	CL:0002397	CD14-positive, CD16-positive monocyte

12. Lymphocytes of B lineage

Q-value	ID	Term
3.28E-10	CL:0000945	lymphocyte of B lineage
2.30E-07	CL:0000542	lymphocyte
2.30E-07	CL:0002242	nucleate cell
2.32E-06	CL:0002087	nongranular leukocyte
3.32E-05	CL:0000738	leukocyte
3.94E-04	CL:0000988	hematopoietic cell
6.39E-04	CL:0000219	motile cell
8.02E-02	CL:0000236	B cell
1.58E-03	DOID:0060058	lymphoma
1.09E-02	DOID:0060083	immune system cancer
1.09E-02	DOID:2531	hematologic cancer

13. Lymphoma

Q-value	ID	Term
1.79E-08	CL:0000542	lymphocyte
1.79E-08	CL:0002242	nucleate cell
4.77E-08	CL:0000084	T cell
1.10E-07	CL:0000738	leukocyte
2.20E-07	CL:0002087	nongranular leukocyte
2.70E-06	CL:0000988	hematopoietic cell
5.47E-06	CL:0000219	motile cell
2.09E-09	DOID:0060083	immune system cancer
2.09E-09	DOID:2531	hematologic cancer
4.75E-05	DOID:0050686	organ system cancer
7.01E-05	DOID:0060058	lymphoma
3.19E-04	DOID:4	disease
2.40E-03	DOID:1240	leukemia
2.40E-03	DOID:162	cancer
2.60E-03	DOID:14566	disease of cellular proliferation

14. Immune organs

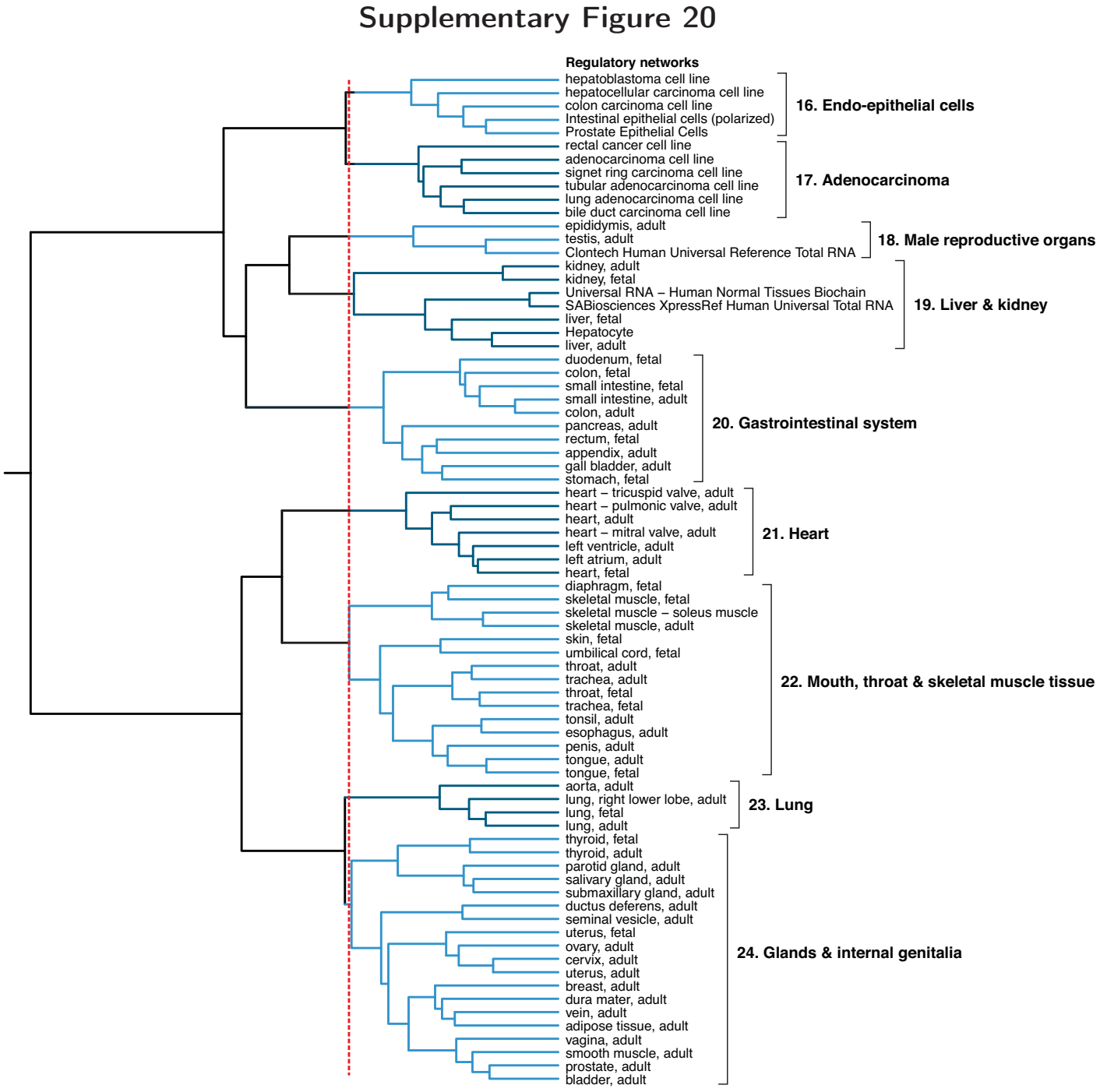
Q-value	ID	Term
8.51E-05	UBERON:0002193	hemolymphoid system
1.11E-04	UBERON:0004177	hemopoietic organ
1.11E-04	UBERON:0005057	immune organ
5.96E-04	UBERON:0002390	hematopoietic system
5.96E-04	UBERON:0002405	immune system
2.60E-02	UBERON:0000077	mixed endoderm/mesoderm-derived structure
3.13E-02	UBERON:0002370	thymus
3.13E-02	UBERON:0005058	hemolymphoid system gland
3.13E-02	UBERON:0009113	thymic region
7.58E-02	UBERON:0002106	spleen

15. Myeloid leukemia

Q-value	ID	Term
6.12E-13	CL:0000763	myeloid cell
7.02E-08	CL:0000988	hematopoietic cell
3.25E-17	DOID:8692	myeloid leukemia
1.92E-13	DOID:1240	leukemia
2.48E-10	DOID:0060083	immune system cancer
2.48E-10	DOID:2531	hematologic cancer
4.46E-05	DOID:0050686	organ system cancer
1.29E-03	DOID:4	disease
3.73E-03	DOID:162	cancer
4.39E-03	DOID:14566	disease of cellular proliferation

Supplementary Figure 19: Dendrogram and annotation of immune system clusters Explanation in legend of **Supplementary Fig. 16**.

Nature Methods: doi:10.1038/nmeth.3799



Supplementary Figure 20: Dendrogram of trunk organs clusters

Dendrogram showing networks and cluster names for (iv) Trunk organs (cf. **Supplementary Fig. 15**). Corresponding annotations are shown in **Supplementary Fig. 21**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 21

16. Endo-epithelial cells

Q-value	ID	Term
2.79E-02	UBERON:0001242	intestinal mucosa
2.79E-02	UBERON:0001262	wall of intestine
2.79E-02	UBERON:0001277	intestinal epithelium
2.79E-02	UBERON:0004786	gastrointestinal system mucosa
2.79E-02	UBERON:0004808	gastrointestinal system epithelium
3.98E-02	UBERON:0004119	endoderm-derived structure
4.60E-02	UBERON:0000485	simple columnar epithelium
4.60E-02	UBERON:0003929	gut epithelium
7.21E-02	UBERON:0003350	epithelium of mucosa
6.13E-02	UBERON:0001007	digestive system
2.37E-02	CL:0002563	intestinal epithelial cell
5.22E-02	CL:0002076	endo-epithelial cell
6.13E-02	CL:0000066	epithelial cell
8.44E-02	CL:0000181	metabolising cell
8.44E-02	CL:0000182	hepatocyte
8.44E-02	CL:0000412	polyploid cell
8.44E-02	CL:0000417	endopolyploid cell

17. Adenocarcinoma

Q-value	ID	Term
1.34E-03	DOID:299	adenocarcinoma
8.63E-03	DOID:305	carcinoma
2.38E-02	DOID:162	cancer
2.59E-02	DOID:14566	disease of cellular proliferation
3.35E-02	DOID:0050687	cell type cancer
3.35E-02	DOID:4	disease

18. Male reproductive organs

Q-value	ID	Term
4.81E-02	UBERON:0003135	male reproductive organ

19. Liver & kidney

Q-value	ID	Term
3.68E-03	UBERON:0005172	abdomen organ
3.68E-03	UBERON:0005173	abdominal segment organ
4.96E-03	UBERON:0000916	abdomen
4.96E-03	UBERON:0002417	abdominal segment of trunk
3.52E-02	UBERON:0002107	liver
3.52E-02	UBERON:0006925	digestive gland
6.17E-02	UBERON:0005177	trunk organ
6.61E-02	UBERON:0009569	subdivision of trunk
7.75E-02	UBERON:0002423	hepatobiliary system

20. Gastrointestinal system

Q-value	ID	Term
1.76E-06	UBERON:0005409	gastrointestinal system
1.76E-05	UBERON:0000160	intestine
4.06E-05	UBERON:0004921	subdivision of digestive tract
2.47E-04	UBERON:0001007	digestive system
2.52E-04	UBERON:0001555	digestive tract
9.30E-04	UBERON:0000059	large intestine
2.71E-03	UBERON:0000481	multi-tissue structure
1.07E-02	UBERON:0001155	colon
1.69E-02	UBERON:0004119	endoderm-derived structure
3.61E-02	UBERON:0000922	embryo
8.51E-03	UBERON:0011216	organ system subdivision

21. Heart

Q-value	ID	Term
6.86E-10	UBERON:0007100	circulatory organ
1.33E-07	UBERON:0000948	heart
3.14E-06	UBERON:0003103	compound organ
2.07E-05	UBERON:0001009	circulatory system
4.73E-04	UBERON:0000946	cardial valve
4.73E-04	UBERON:0003978	valve
4.73E-04	UBERON:0004151	cardiac chamber
6.52E-04	UBERON:0004535	cardiovascular system
1.56E-02	UBERON:0010313	neural crest-derived structure
2.09E-02	UBERON:0010314	structure with developmental contribution from neural crest
2.16E-02	UBERON:0002081	cardiac atrium
2.16E-02	UBERON:0002133	atrioventricular valve
3.38E-02	UBERON:0007023	adult organism

22. Mouth, throat & skeletal muscle tissue

Q-value	ID	Term
1.07E-02	UBERON:0000475	organism subdivision
7.49E-02	UBERON:0000341	throat
8.34E-03	UBERON:0001134	skeletal muscle tissue
8.34E-03	UBERON:0002036	striated muscle tissue
3.61E-02	UBERON:0000383	musculature of body
3.61E-02	UBERON:0001015	musculature
3.61E-02	UBERON:0002385	muscle tissue
7.49E-02	UBERON:0001630	muscle organ
7.49E-02	UBERON:0005090	muscle structure
6.12E-03	UBERON:0000922	embryo

23. Lung

Q-value	ID	Term
7.98E-03	UBERON:0000025	tube
1.23E-02	UBERON:0000117	respiratory tube
1.23E-02	UBERON:0000170	pair of lungs
1.23E-02	UBERON:0000171	respiration organ
1.23E-02	UBERON:0002048	lung
3.02E-02	UBERON:0004111	anatomical conduit
3.38E-02	UBERON:0000065	respiratory tract
3.68E-02	UBERON:0005178	thoracic cavity organ
8.77E-02	UBERON:0001004	respiratory system
4.17E-02	UBERON:0005181	thoracic segment organ
8.15E-02	UBERON:0000915	thoracic segment of trunk

24. Glands & internal genitalia

Q-value	ID	Term
1.87E-02	UBERON:0004175	internal genitalia
2.39E-02	UBERON:0000990	reproductive system
2.39E-02	UBERON:0005156	reproductive structure
3.57E-02	UBERON:0003133	reproductive organ
2.76E-02	UBERON:0001044	salivary gland
2.76E-02	UBERON:0003294	gland of foregut
2.76E-02	UBERON:0003408	gland of gut
2.76E-02	UBERON:0010047	oral gland
3.59E-02	UBERON:0002530	gland
8.27E-02	UBERON:0002553	anatomical cavity
7.00E-08	UBERON:0007023	adult organism

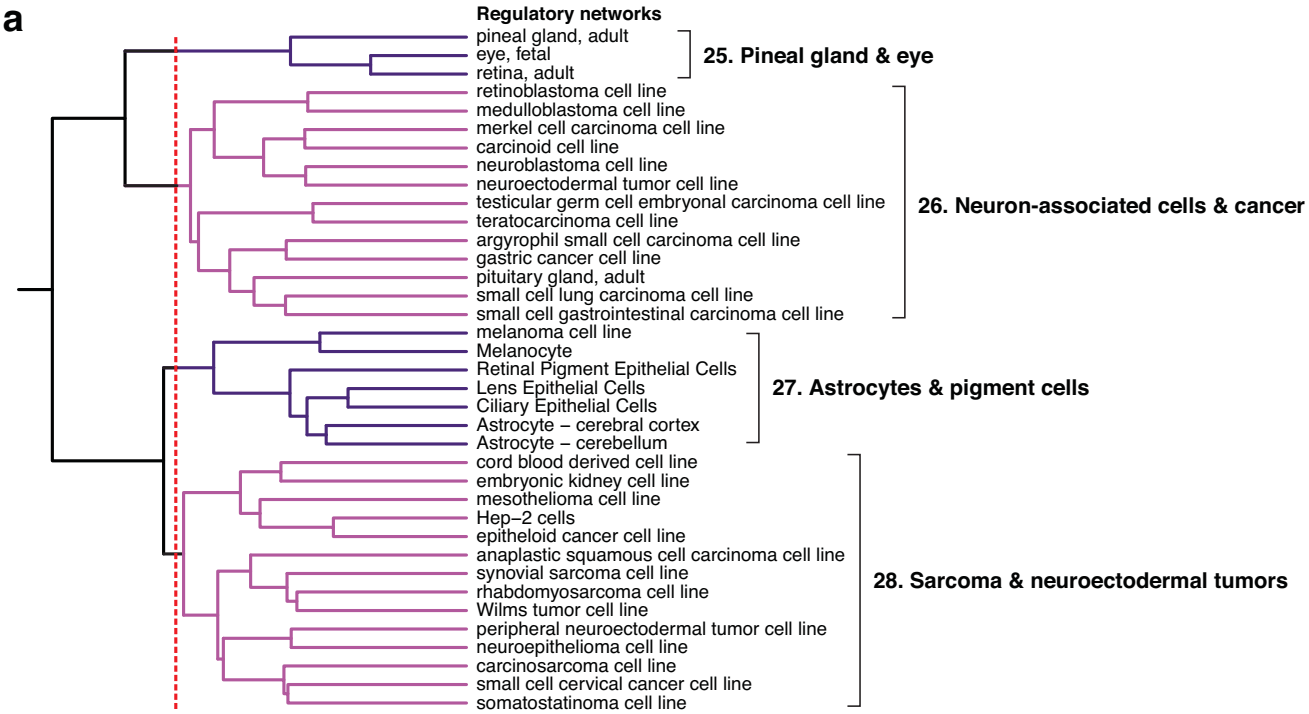
Supplementary Figure 21: Annotation of trunk organs clusters

Enriched annotations for each cluster shown in **Supplementary Fig. 20**. Explanation in legend of **Supplementary Fig. 16**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 22

a



b

25. Eye

Q-value	ID	Term
No enriched terms		

26. Neuron-associated cells & cancer

Q-value	ID	Term
9.22E-02	CL:0000095	neuron associated cell
3.61E-04	DOID:162	cancer
4.21E-04	DOID:14566	disease of cellular proliferation
5.19E-04	DOID:0050687	cell type cancer
7.57E-04	DOID:4	disease
5.83E-03	DOID:2994	germ cell cancer
5.83E-03	DOID:3095	germ cell and embryonal cancer
1.89E-02	DOID:305	carcinoma
2.63E-02	DOID:0050686	organ system cancer
4.15E-02	DOID:3119	gastrointestinal system cancer

27. Astrocytes & pigment cells

Q-value	ID	Term
3.67E-03	UBERON:0007625	pigment epithelium of eye
4.67E-03	UBERON:0010371	ecto-epithelium
2.79E-02	UBERON:0000019	camera-type eye
2.79E-02	UBERON:0000047	simple eye
2.79E-02	UBERON:0004088	ocular region
2.79E-02	UBERON:0010230	eyeball of camera-type eye
3.52E-02	UBERON:0000970	eye
4.33E-02	UBERON:0001456	face
4.33E-02	UBERON:0002104	visual system
5.10E-02	UBERON:0004121	ectoderm-derived structure
6.61E-02	UBERON:0000020	sense organ
7.75E-02	UBERON:0001032	sensory system
7.75E-02	UBERON:0004456	entire sense organ system
1.28E-02	CL:0000075	columnar/cuboidal epithelial cell
4.45E-02	CL:0000126	macroglial cell
4.45E-02	CL:0000127	astrocyte
4.45E-02	CL:0000128	oligodendrocyte
4.45E-02	CL:0000147	pigment cell
6.88E-02	CL:0000325	stuff accumulating cell

28. Sarcoma & neuroectodermal tumors

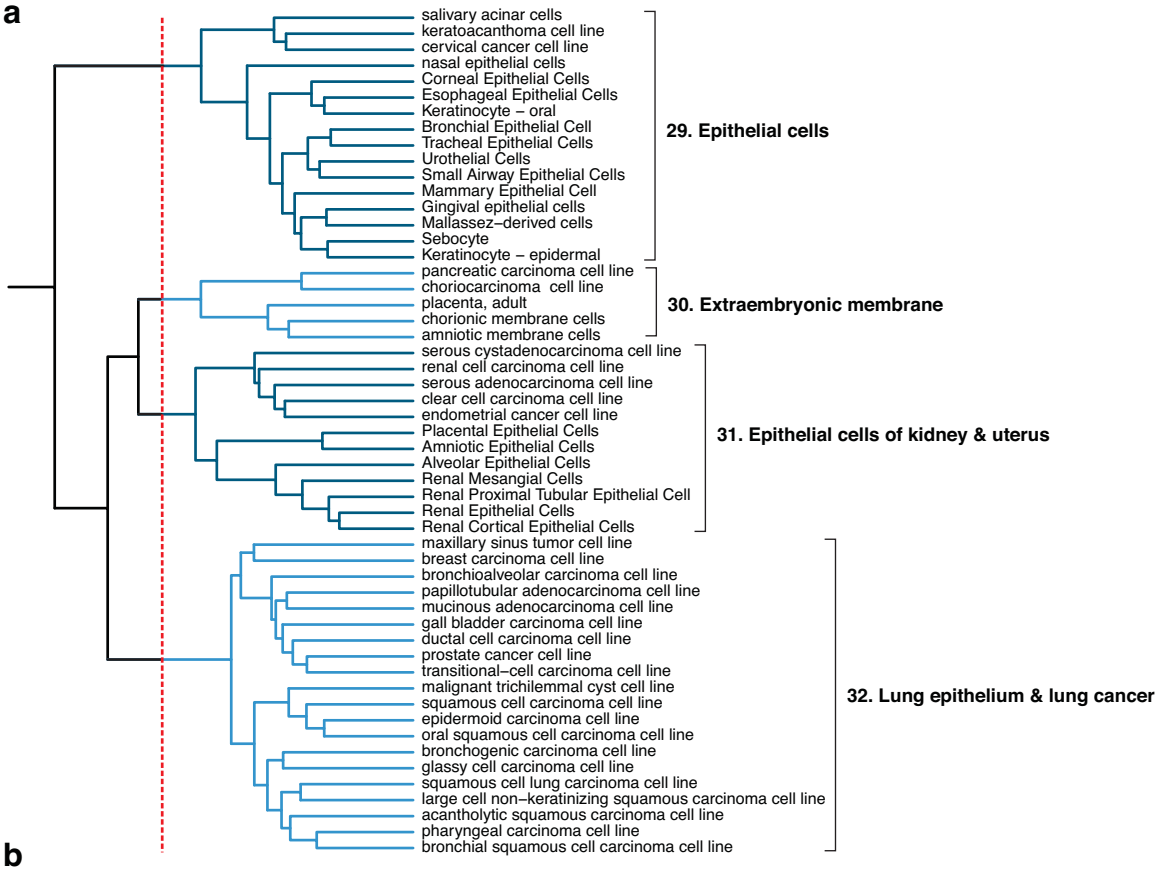
Q-value	ID	Term
1.34E-03	DOID:0050687	cell type cancer
2.40E-03	DOID:1115	sarcoma
2.60E-03	DOID:4	disease
8.63E-03	DOID:162	cancer
1.00E-02	DOID:14566	disease of cellular proliferation
1.21E-02	DOID:171	neuroectodermal tumor
9.65E-02	DOID:169	neuroendocrine tumor

Supplementary Figure 22: Dendrogram and annotation of neuron-associated clusters
Explanation in legend of **Supplementary Fig. 16**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 23

a



b

29. Epithelial cells

Q-value	ID	Term
4.89E-05	CL:0000066	epithelial cell
2.75E-04	CL:0002076	endo-epithelial cell
1.78E-03	CL:0002077	ecto-epithelial cell
2.03E-03	CL:0002159	general ecto-epithelial cell
2.98E-02	CL:0002251	epithelial cell of alimentary canal
5.13E-02	CL:0002368	respiratory epithelial cell
7.29E-02	CL:0000622	acinar cell

30. Extraembryonic membrane

Q-value	ID	Term
1.37E-02	UBERON:0000478	extraembryonic structure
4.60E-02	UBERON:0000158	membranous layer
4.60E-02	UBERON:0005631	extraembryonic membrane

32. Lung epithelium & lung cancer

Q-value	ID	Term
7.21E-02	UBERON:0000115	lung epithelium
2.58E-02	UBERON:0000464	anatomical space
2.58E-02	UBERON:0000466	immaterial anatomical entity
7.21E-02	UBERON:0004802	respiratory tract epithelium
9.57E-02	UBERON:0004807	respiratory system epithelium
9.73E-02	UBERON:0005911	endo-epithelium
1.94E-08	CL:0000066	epithelial cell
4.48E-04	CL:0000076	squamous epithelial cell
6.20E-02	CL:0000082	epithelial cell of lung
2.40E-03	DOID:0050615	respiratory system cancer
1.80E-02	DOID:1324	lung cancer
2.70E-14	DOID:305	carcinoma
1.99E-11	DOID:0050687	cell type cancer
2.82E-09	DOID:162	cancer
3.63E-09	DOID:14566	disease of cellular proliferation
1.11E-08	DOID:4	disease
1.21E-08	DOID:1749	squamous cell carcinoma

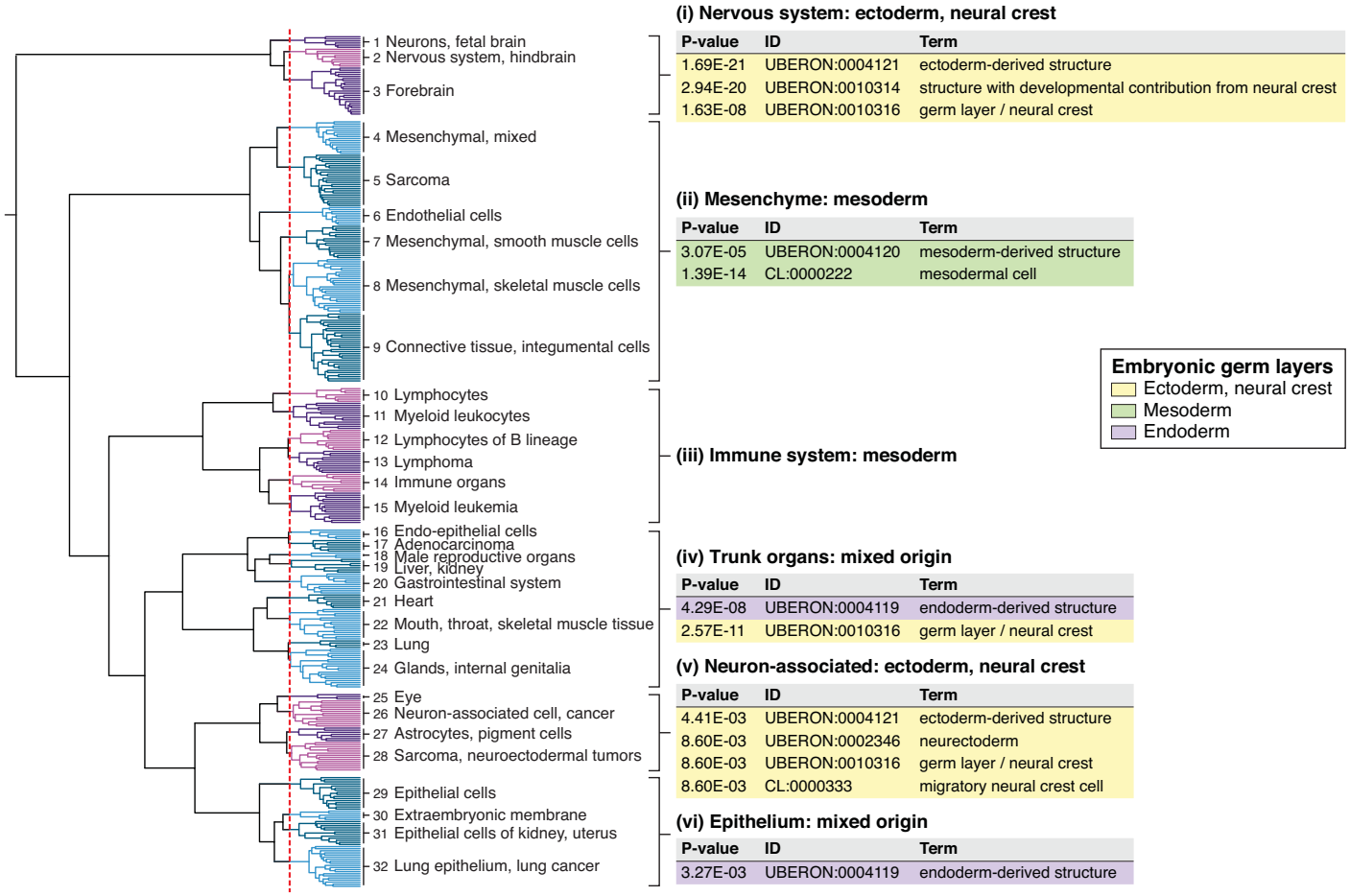
31. Epithelial cells of kidney & uterus

Q-value	ID	Term
1.61E-05	UBERON:0001285	nephron
1.61E-05	UBERON:0004819	kidney epithelium
1.61E-05	UBERON:0006555	excretory tube
1.61E-05	UBERON:0007684	uriniferous tubule
3.59E-05	UBERON:0002113	kidney
3.59E-05	UBERON:0011143	upper urinary tract
9.47E-05	UBERON:0001231	nephron tubule
9.47E-05	UBERON:0004211	nephron epithelium
9.47E-05	UBERON:0004810	nephron tubule epithelium
9.47E-05	UBERON:0009773	renal tubule
1.41E-04	UBERON:0001008	excretory system
1.41E-04	UBERON:0006554	urinary system structure
2.20E-04	UBERON:0000489	cavitated compound organ
3.58E-04	UBERON:0001225	cortex of kidney
3.58E-04	UBERON:0008987	renal parenchyma
1.95E-03	UBERON:0000353	parenchyma
1.95E-03	UBERON:0001851	cortex
7.64E-03	UBERON:0005172	abdomen organ
7.64E-03	UBERON:0005173	abdominal segment organ
9.02E-03	UBERON:0003103	compound organ
1.07E-02	UBERON:0000916	abdomen
1.07E-02	UBERON:0002417	abdominal segment of trunk
4.33E-02	UBERON:0003914	epithelial tube
9.57E-02	UBERON:0000474	female reproductive system
2.66E-07	CL:0002518	kidney epithelial cell
2.66E-07	CL:1000497	kidney cell
9.71E-06	CL:1000449	epithelial cell of nephron
1.22E-05	CL:0000066	epithelial cell
6.67E-05	CL:1000494	epithelial cell of renal tubule
6.67E-05	CL:1000507	kidney tubule cell
2.75E-04	CL:0002584	renal cortical epithelial cell
2.75E-04	CL:0002681	kidney cortical cell
4.60E-02	CL:0002149	epithelial cell of uterus
4.60E-02	CL:0002255	stromal cell of endometrium
6.16E-03	DOID:120	female reproductive organ cancer
1.11E-02	DOID:193	reproductive organ cancer
1.89E-02	DOID:299	adenocarcinoma

Supplementary Figure 23: Dendrogram and annotation of epithelium clusters
Explanation in legend of **Supplementary Fig. 16**.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 24

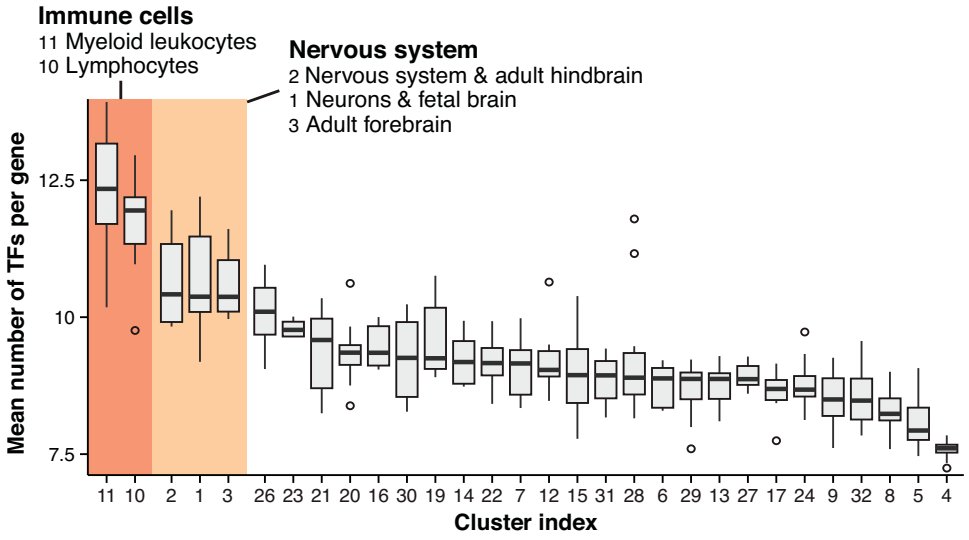


Supplementary Figure 24: Developmental origin of regulatory networks pertaining to different high-level clusters

We performed annotation enrichment analysis for each high-level cluster (i-vi, **Supplementary Fig. 15**), focusing specifically on terms related to the three embryonic germ layers (ectoderm and neural crest, mesoderm, and endoderm). Regulatory networks of the first cluster (i, nervous system) are derived from ectoderm and neural crest. Regulatory networks of the second cluster (ii, mesenchyme) are mesoderm-derived. Immune cells (clusters 11-13) are derived from hematopoietic stem cells (the corresponding high-level cluster [iii, immune system] does not show significant enrichment for mesoderm; this is just an artifact because immune cells and corresponding cancer cells were not systematically annotated as mesodermal cells in the FANTOM5 sample ontology). The fourth cluster (iv, trunk organs) groups regulatory networks of mixed origin (endoderm and neural crest show significant enrichment, but we also found mesoderm-derived tissues). The fifth cluster (v, neuron-associated) enriches for ectoderm and neural crest. Epithelial cells (cluster vi) are derived from all three germ layers. In summary, regulatory networks that are clustered together often share a common developmental origin.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 25

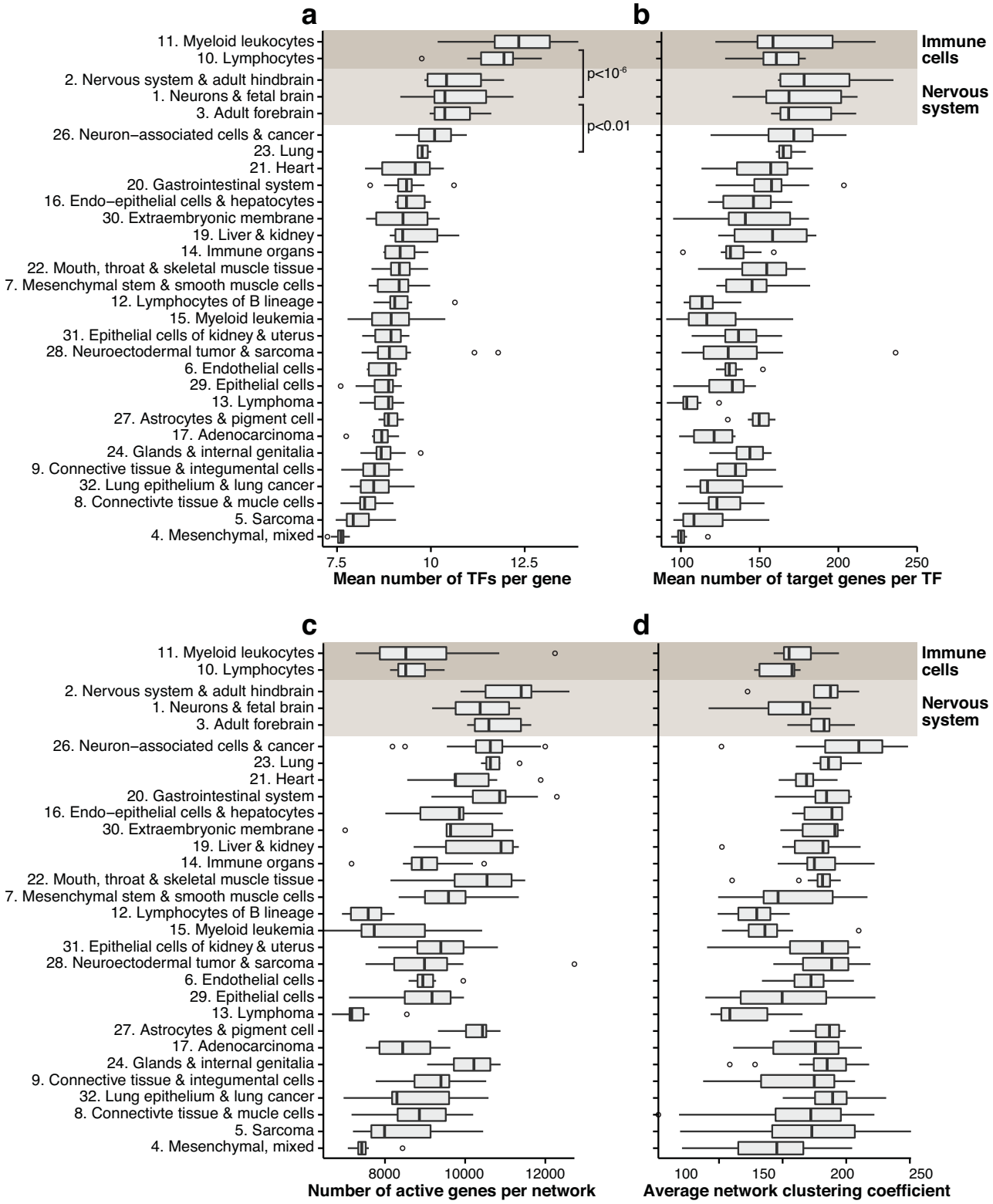


Supplementary Figure 25: Comparing number of TFs per gene across lineages

Boxplots show the mean number of TF inputs per gene (i.e., the mean indegree) for all networks across the 32 clusters defined in **Supplementary Fig. 14** (weak edges with weights below 0.05 were not counted for the indegrees). Immune cells, followed by cells and tissues of the nervous system, have the highest number of TFs per gene, indicating that they rely on more intricate, combinatorial regulation than other cells and tissues (the observed differences between the mean indegrees of immune cells, nervous system, and the remaining clusters are statistically significant; **Supplementary Fig. 26**). Both immune cells and neurons perform highly adaptive functions that require integration of diverse extra-cellular signals, e.g., to recognize distinct pathogens or guide neuronal wiring. Our networks suggest that the orchestration of transcriptional responses in these cells involves intricate regulatory programs.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 26



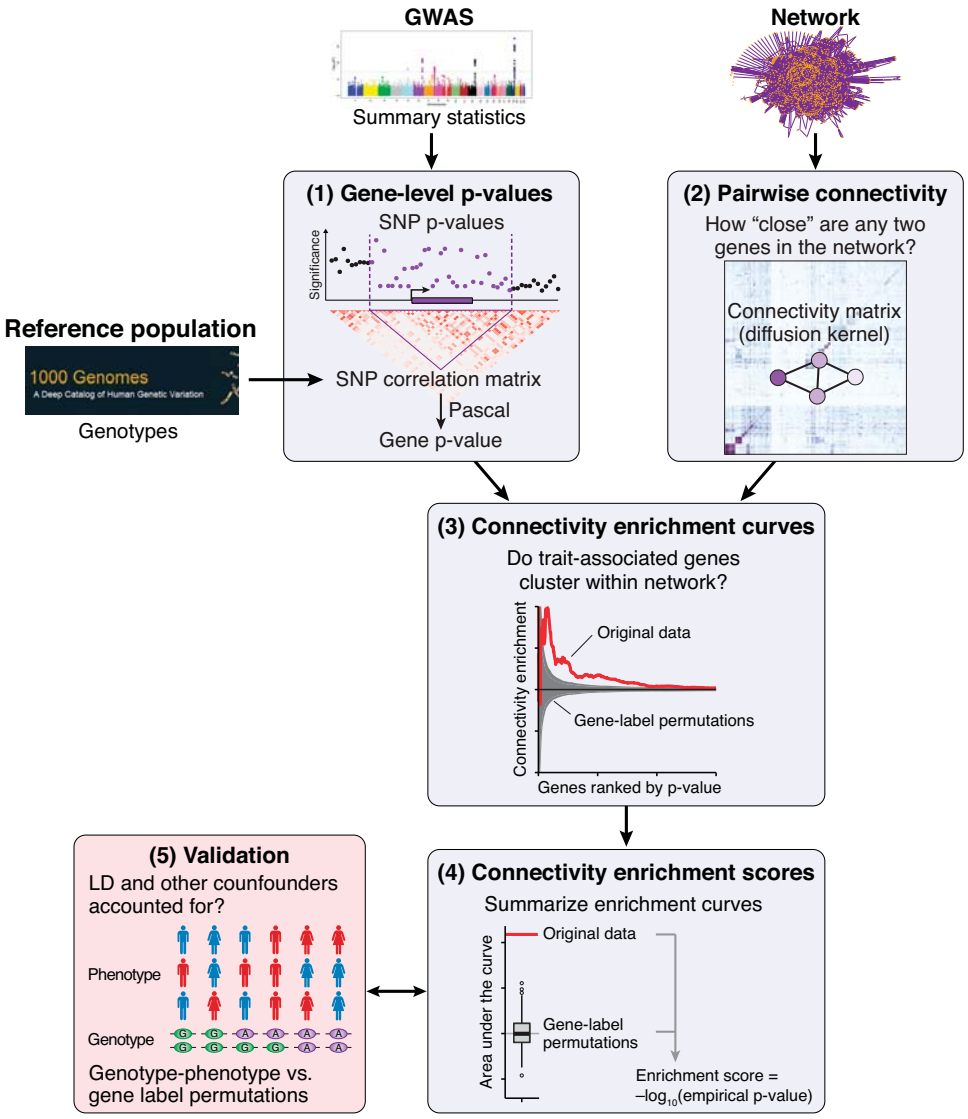
Supplementary Figure 26: Comparing network properties across lineages
(Caption next page.)

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 26: (Previous page.)
Boxplots show network properties for all 394 regulatory networks, grouped according to the 32 clusters defined in **Supplementary Fig. 14**. The same cluster order on the vertical axis is used in all four panels.
(**a**) Mean number of TFs per gene (i.e., mean indegree; same as **Supplementary Fig. 25**). Clusters were ordered on the vertical axis by their median. Immune cells, followed by cells and tissues of the nervous system, have the highest number of TFs per gene. The observed differences are statistically significant: immune cells have a higher mean indegree than nervous system networks ($p<10\text{-}6$, two-sided Wilcoxon rank-sum test), and the latter come before other networks with a high mean indegree, such as those of lung ($p<0.01$). Note, cluster 26 (neuron-associated cells and cancer) also contains some neural tissues and related tumors, which explains why it comes right after the nervous system clusters.
(**b**) Mean number of target genes per TF (i.e., mean outdegree). Networks of immune cells and nervous system clusters have relatively high mean outdegree; however, in contrast to the mean indegree, differences between immune cells, nervous system, and lung (as an example of another cluster with high mean outdegree) are not statistically significant ($p>0.05$).
(**c**) The number of active genes in each network mainly depends on whether it is derived from a cell or tissue sample. Tissues comprise multiple cell types and thus tend to have more active genes (the union of the genes active in each cell type). Nervous system networks have a similar number of active genes as other tissue-specific networks (e.g., lung). Immune cell networks have a lower number of active genes, but again not significantly different from other cell type-specific networks (e.g., connective tissue and muscle cells).
(**d**) There are no notable differences in the average network clustering coefficient between different lineages.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 27



Supplementary Figure 27: Overview of network connectivity enrichment analysis

Given summary statistics from a GWAS and a network, our pipeline evaluates whether genes perturbed by trait-associated variants are more densely inter-connected than expected. Here we give a general overview of the approach, see **Supplementary Fig. 28** for an illustrative example.

(1) SNP association p-values from the GWAS are summarized at the level of genes using Pascal,[8] a fast and accurate tool that corrects for LD structure using information from a reference population (we used the European panel of the 1000 genomes project[9]).

(2) A pairwise connectivity matrix is computed, which defines how "close" any two genes are in the network based on a random-walk kernel.[10]

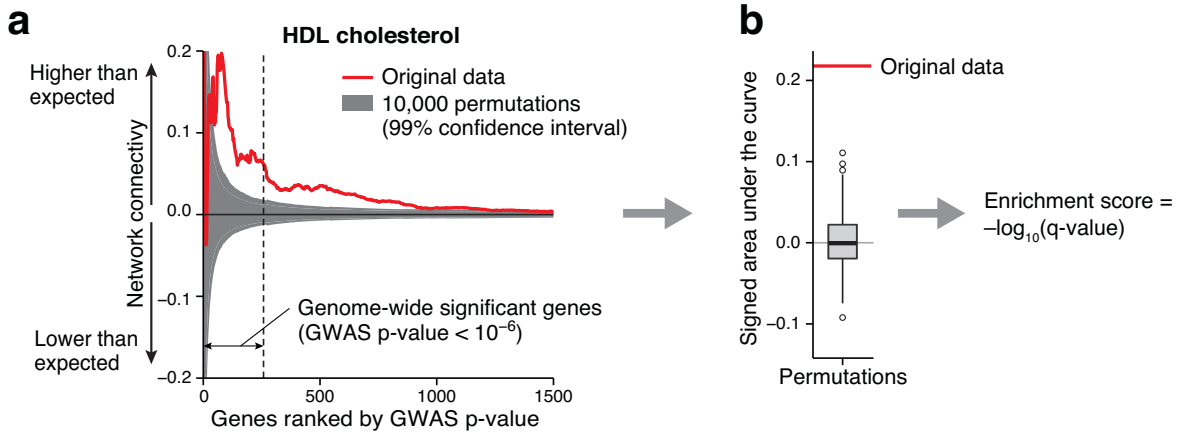
(3) Genes are ranked by their GWAS p-value, from most to least significant, and the mean connectivity between the top k genes is evaluated — as k is varied over the complete list — both for the original data and random permutations (**Supplementary Fig. 28**). A conservative approach is used for the permutations, where the network structure remains completely fixed and only the labels of genes with similar network degree are shuffled.[11] Gene pairs that are in LD are excluded from the analysis (**Supplementary Fig. 29**).

(4) Enrichment curves are summarized by the signed area under the curve, both for the original data and the permutations, and a corresponding empirical p-value is computed.

(5) We validated that LD structure and other potential confounders are accounted for by showing that enrichment scores based on gene label permutations are equivalent to those obtained using genotype-phenotype permutations (**Supplementary Fig. 30**).

Nature Methods: doi:10.1038/nmeth.3799

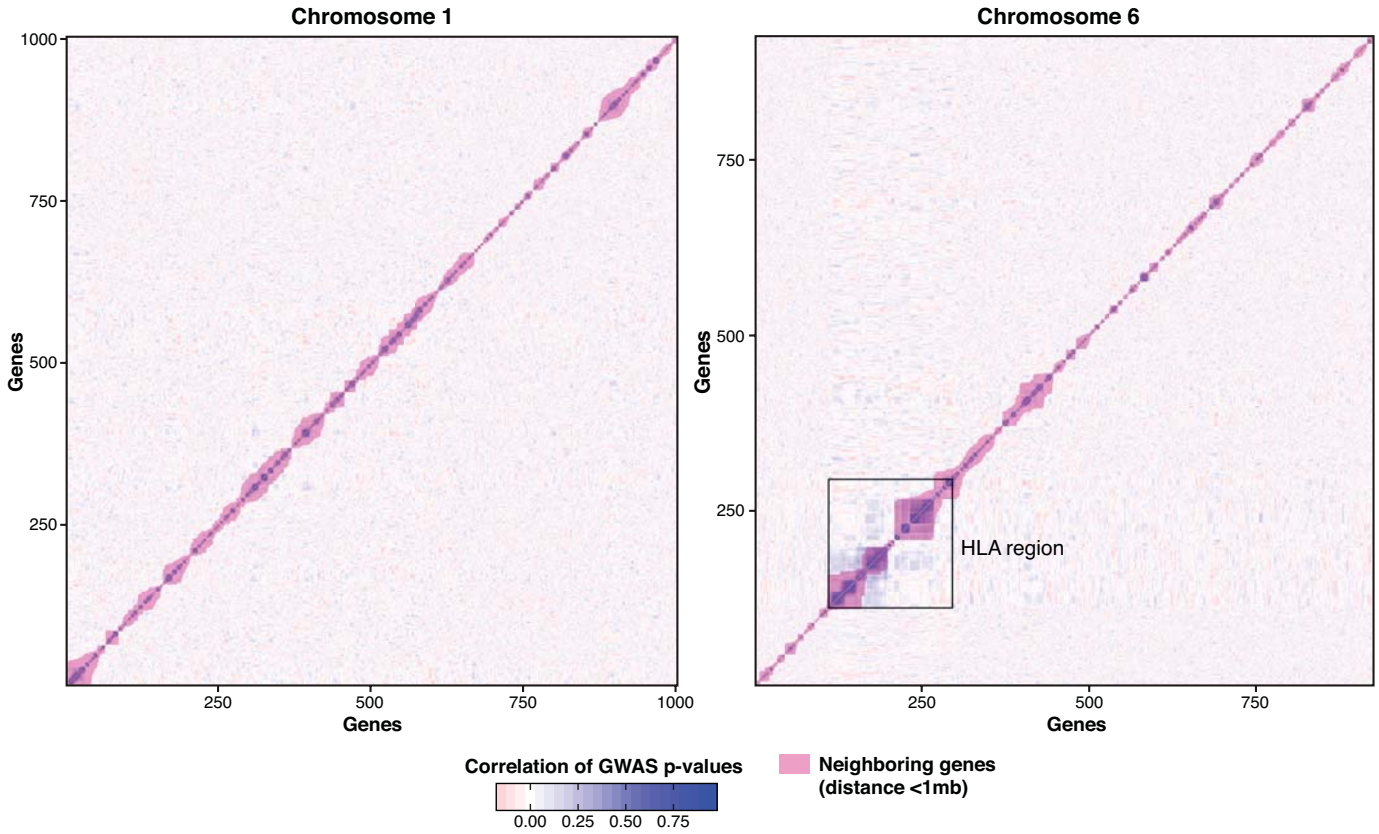
Supplementary Figure 28



Supplementary Figure 28: Illustrative example for network connectivity enrichment analysis
Network connectivity analysis is illustrated using an example trait and network that show strong enrichment (HDL cholesterol and the global regulatory network defined in Methods).
(**a**) Genes are ranked by GWAS p-value. For each position k in the ranked list, the network connectivity between the top k genes is computed using a random-walk kernel (red line; positive and negative values indicate higher and lower connectivity than expected, respectively). The same is done for 10,000 permutations of the data (the grey area contains 99% of the corresponding curves). In this example, the top 500 genes show significantly increased connectivity, which includes genes with GWAS p-values that do not reach the genome-wide significance threshold (dashed line).
(**b**) To summarize connectivity enrichment, the signed area under the curve is computed both for the original data (red line) and the permutations (boxplot), resulting in an empirical p-value and corresponding score that gauge how clustered trait-associated genes are within the network.

Nature Methods: doi:10.1038/nmeth.3799

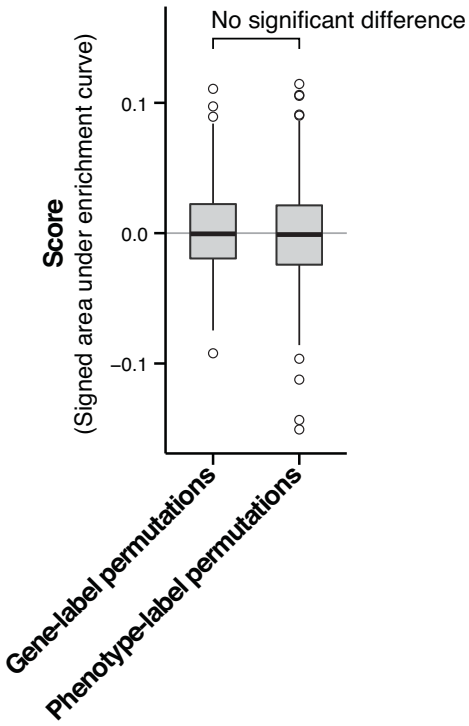
Supplementary Figure 29



Supplementary Figure 29: Correlation between neighboring genes due to linkage disequilibrium
The GWAS association signal of neighboring genes is often correlated due to linkage disequilibrium (LD). Here we show the pairwise correlation between gene scores (GWAS p-values summarized by gene using Pascal[8]) across 500 simulated phenotypes (Methods). Genes of chromosome 1 and 6 are shown, ordered by genomic position (the same observations were be made for other chromosomes). Genes that are less than 1 mega-base apart (magenta overlay) often show increased correlation. This is relevant for network analysis because neighboring genes are often also functionally related and/or co-regulated, i.e., they are also neighbors at the level of pathways or networks. An extreme example is the human leukocyte antigen (HLA) gene cluster on chromosome 6. If not properly corrected for, such clusters of genes that are both correlated and functionally related lead to inflated pathway or network enrichment scores. To ensure that our results were not confounded by neighboring, functionally related genes, we took a conservative approach and completely excluded all gene pairs with a distance of less than 1 mega-base from the network connectivity enrichment analysis (Methods). Since the HLA genes form an exceptionally large cluster that also shows strong association with many immune-related traits, we further completely excluded all genes in the HLA region from the connectivity analysis. This ensures that the observed network connectivity enrichment is not driven by the HLA genes or similar gene clusters.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 30

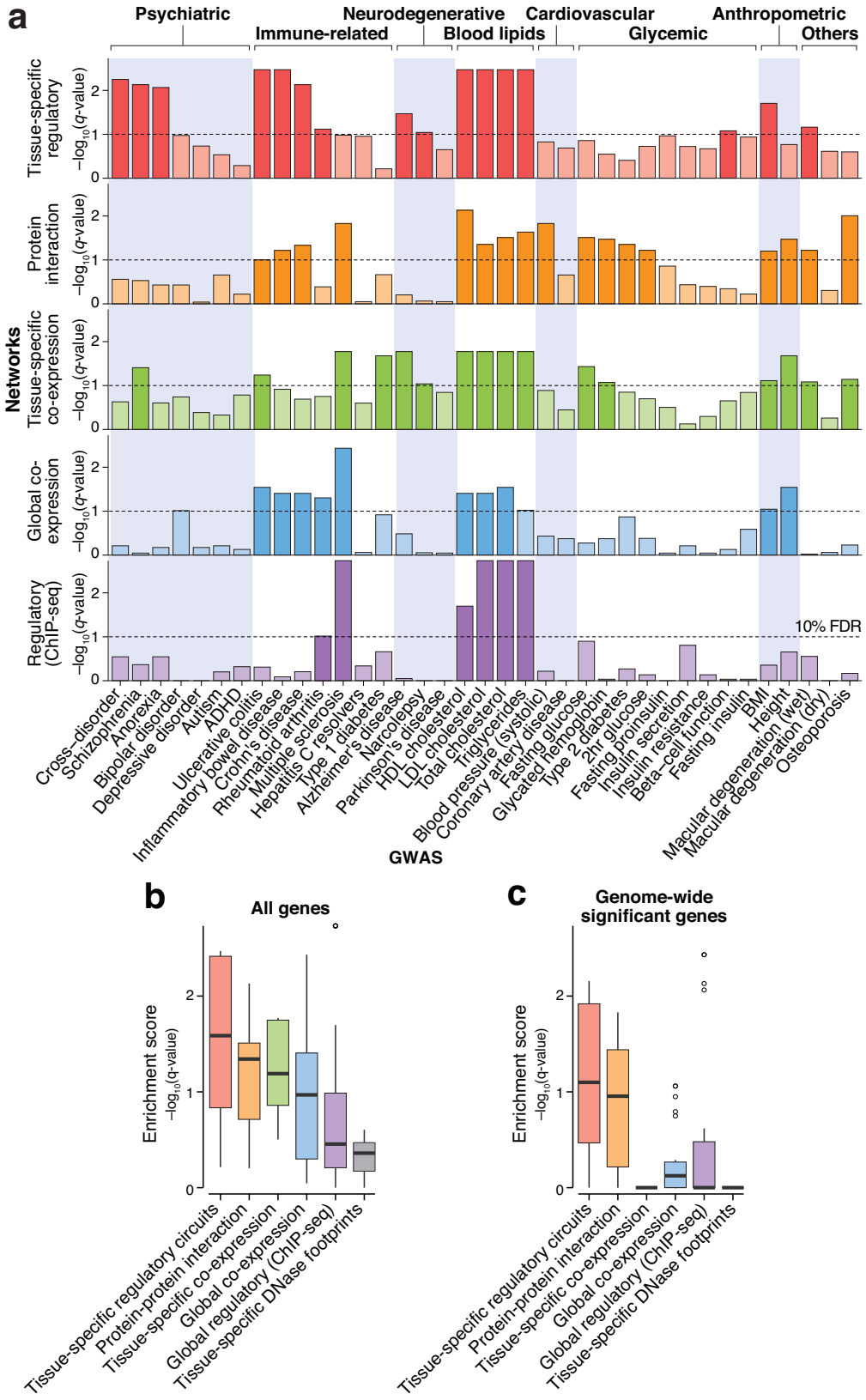


Supplementary Figure 30: Validation of connectivity enrichment scores using phenotype-label permutations

Our method computes connectivity enrichment scores empirically based on permutations of the data. It is thus crucial that permutations are done appropriately such that no bias is introduced. Phenotype label permutation is generally considered the gold standard for this purpose because it preserves both the network structure and genomic architecture (e.g., LD structure and gene clusters, see **Supplementary Fig. 29**). However, it is computationally intensive and requires access to genotypic data, which are rarely shared. Thus, a standard approach used in network-based GWAS analysis is within-degree gene label permutation: the labels of genes with similar network degree are shuffled, while the network structure itself remains completely fixed[11,12] (i.e., no edges are rewired, only node labels are reassigned). However, in our case the relevant confounder is not the degree of the gene, but rather its mean pairwise connectivity across all other genes (we call this the *centrality* of the gene), which is closely related — but not equivalent — to the degree. Thus, we used the same approach as within-degree gene label permutation, but in our case the genes were binned based on their centrality (Methods). This permutation method implies that the null distribution of connectivity between the top k genes is entirely a function of the centrality of these genes.[11] In order to validate the within-centrality gene label permutation method, we generated 500 phenotype-label permutations and compared the resulting scores (signed area under the connectivity enrichment curve; Methods). We found no significant difference ($p=0.18$, two-sided Wilcoxon rank-sum test), thus confirming that our method based on within-centrality gene label permutation adequately corrects for confounders. I.e., the resulting enrichment scores are equivalent to those obtained using phenotype label permutation. Note that this only works because our method excludes gene pairs that are in LD as explained in **Supplementary Fig. 29** — if these gene pairs are included scores become inflated.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 31



Supplementary Figure 31: Connectivity enrichment scores for different types of networks (Caption next page.)

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 31: (Previous page.)
Connectivity enrichment scores for different types of networks and GWAS traits. See legend of Fig. 2c and Methods for a description of the different networks.
(**a**) Same as Fig. 2c, but here we show results for all traits, not just those showing significant connectivity enrichment for at least one network. Some traits do not show enrichment, which may be either because the signal was too weak, the right tissues were not profiled, or other types of networks may be more relevant for these traits.
(**b-c**) Overall connectivity enrichment scores when considering (**b**) all genes and (**c**) only genes with GWAS p-values that reach genome-wide significance.
(**b**) This panel summarizes the results of Fig. 2c. Tissue-specific regulatory networks show the strongest connectivity enrichment, followed by protein-protein interaction networks and tissue-specific co-expression networks. Among the four protein-protein interaction networks tested (Methods), the InWeb database[13] showed the best connectivity enrichment, covering the largest number of traits. The regulatory networks from ENCODE based on ChIP-seq[14] and DNaseI footprints[15] only show weak connectivity enrichment for most traits. Note that the networks based on DNaseI footprints were likely not comprehensive enough to reveal perturbed regulatory modules because they are limited to TF genes and promoter-level interactions (Methods).
(**c**) When considering only genes that pass the GWAS significance threshold, the connectivity enrichment scores are much lower, suggesting substantial contribution of weakly associated genes.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 32

a Psychiatric

	High-level networks	Individual networks
Cross-disorder	Suppl. Fig. 33	Suppl. Fig. 33
Schizophrenia	Suppl. Fig. 34	Suppl. Fig. 34
Anorexia nervosa	Suppl. Fig. 35	No enrichment
Bipolar disorder	No enrichment	Suppl. Fig. 36
Depressive disorder	No enrichment	No enrichment
Autism	No enrichment	No enrichment
ADHD	No enrichment	No enrichment

b Immune-related

	High-level networks	Individual networks
Inflammatory bowel disease	Suppl. Fig. 37	Suppl. Fig. 37
Ulcerative colitis	Suppl. Fig. 38	Suppl. Fig. 38
Crohn's disease	Suppl. Fig. 39	Suppl. Fig. 39
Rheumatoid arthritis	Suppl. Fig. 40a	Suppl. Fig. 40a
Multiple sclerosis	No enrichment	Suppl. Fig. 40b
Hepatitis C resolvers	No enrichment	Suppl. Fig. 40c
Type 1 diabetes	No enrichment	No enrichment

c Neurodegenerative

	High-level networks	Individual networks
Alzheimer's disease	Suppl. Fig. 41	Suppl. Fig. 41
Narcolepsy	Suppl. Fig. 42	Suppl. Fig. 42
Parkinson's disease	No enrichment	No enrichment

d Blood lipids

	High-level networks	Individual networks
HDL cholesterol	Suppl. Fig. 43a	Not tested
LDL cholesterol	Suppl. Fig. 43b	Not tested
Total cholesterol	Suppl. Fig. 44a	Not tested
Triglycerides	Suppl. Fig. 44b	Not tested

e Cardiovascular

	High-level networks	Individual networks
Blood pressure (systolic)	No enrichment	No enrichment
Coronary artery disease	No enrichment	No enrichment

f Glycemic

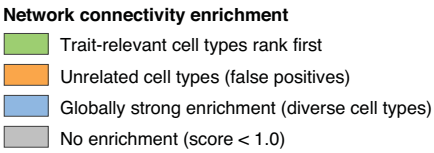
	High-level networks	Individual networks
Fasting glucose	No enrichment	No enrichment
Glycated hemoglobin	No enrichment	No enrichment
Type 2 diabetes	No enrichment	No enrichment
2hr glucose	No enrichment	No enrichment
Fasting proinsulin	No enrichment	No enrichment
Insulin secretion	No enrichment	No enrichment
Insulin resistance	No enrichment	No enrichment
Beta-cell function	Suppl. Fig. 46	No enrichment
Fasting insulin	No enrichment	No enrichment

g Anthropometric

	High-level networks	Individual networks
Body mass index	Suppl. Fig. 45	Suppl. Fig. 45
Height	No enrichment	No enrichment

h Others

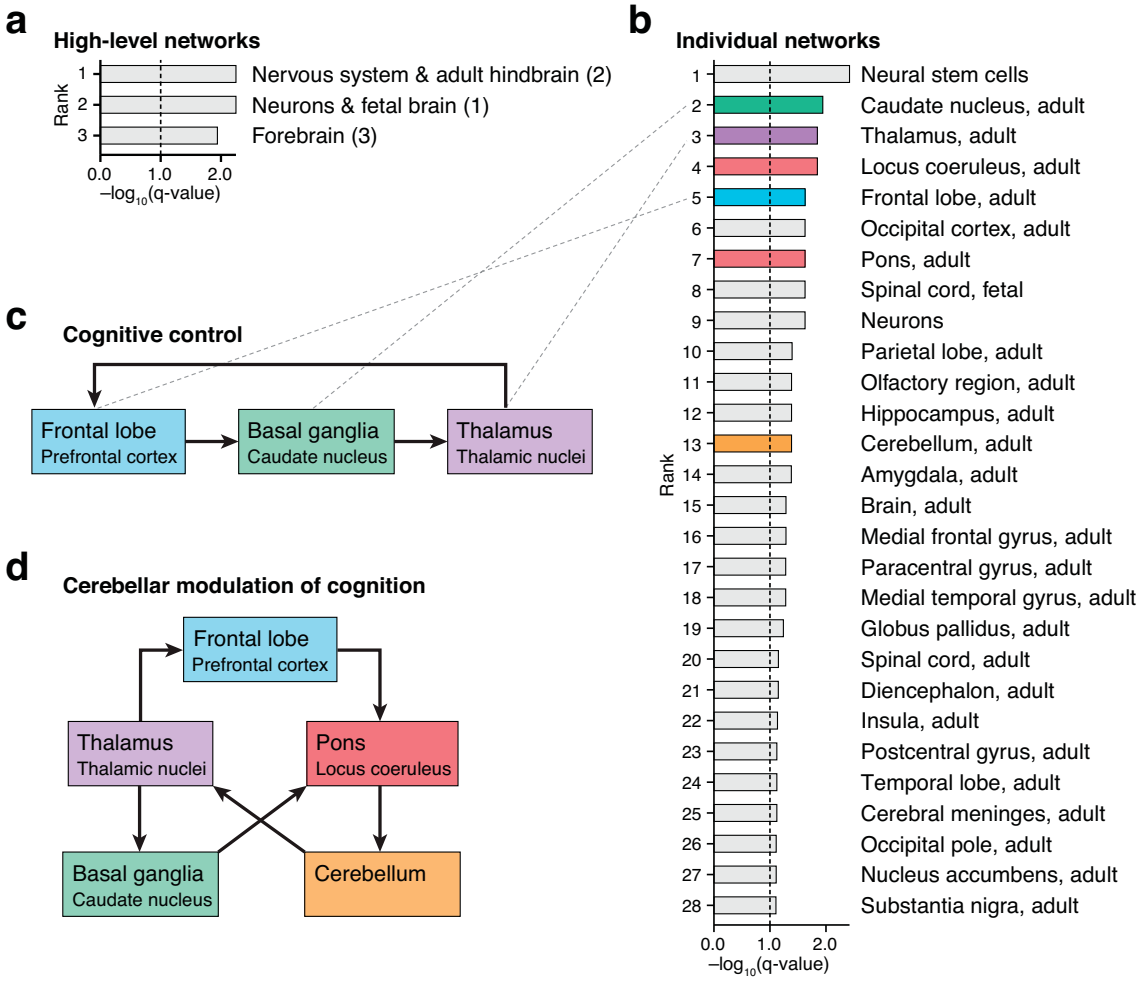
	High-level networks	Individual networks
AMD (neovascular)	Suppl. Fig. 47	Suppl. Fig. 47
AMD (dry)	No enrichment	No enrichment
Osteoporosis	No enrichment	No enrichment



Supplementary Figure 32: Overview of network connectivity enrichment results for all traits
Traits are listed in same order as in **Supplementary Fig. 31**. Refer to the indicated supplementary figures and Results for details. Possible reasons why some traits did not show significant connectivity enrichment are: (1) the signal was too weak (four out of five GWASs with <10,000 individuals were not enriched in any network), (2) the relevant tissues were not profiled (e.g., our library does not include pancreatic islet cells relevant for type 2 diabetes[16]), or (3) other types of networks (e.g., post-transcriptional) may be more relevant for these traits.
(**a**–**c**) Psychiatric, immune-related, and neurodegenerative disorders showed either highly specific connectivity enrichment in disease-related cell types and tissues or no enrichment at all, with the exception of Crohn's disease.
(**d**) Blood lipid traits showed strong connectivity enrichment in many diverse networks.
(**e**–**f**) Cardiovascular and glycemic traits showed no connectivity enrichment except for one likely false positive (β-cell function).
(**g**–**h**) Body mass index and age-related macular degeneration (AMD) of neovascular type consistently showed connectivity enrichment in trait-related networks, while the remaining traits showed no enrichment.

Nature Methods: doi:10.1038/nmeth.3799

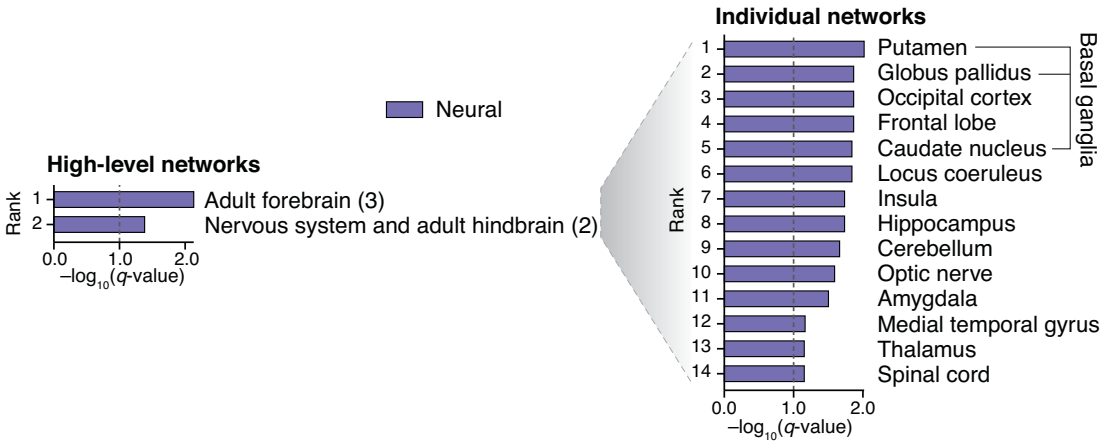
Supplementary Figure 33



Supplementary Figure 33: Connectivity enrichment for the psychiatric, cross-disorder study
Connectivity enrichment scores for the psychiatric, cross-disorder study across (a) the 32 high-level networks and (b) the corresponding individual networks. All networks with score > 1.0 are shown (10% FDR, dashed line). Numbers in parentheses correspond to cluster indexes (**Supplementary Fig. 14**). (**a**) The psychiatric cross-disorder study showed the strongest clustering of perturbed genes precisely in the three high-level networks of nervous system and brain. (**b**) After neural stem cells, the strongest connectivity enrichment was observed in the caudate nucleus, thalamus, and locus coeruleus, which are all implicated in psychiatric disorders. The caudate nucleus is a basic structure of the basal ganglia, strongly innervated by dopamine neurons, which performs important functions related to goal-directed action, memory, learning, and emotion. Dopamine is an important neurotransmitter implicated in psychiatric disorders, e.g., the dopaminergic system of the basal ganglia manifests pathological anomalies in schizophrenia patients and is the primary target of current antipsychotic drugs.[17] The thalamus, located adjacent to the caudate nucleus, is an information processing and communication hub involved in sensory and cognitive processes that has been implicated in the pathophysiology of schizophrenia, for instance.[18] The locus coeruleus is part of the pons, which also showed strong signal (7th rank). It is responsible for physiological responses to emotional pain, stress, and panic and has widespread projections utilizing the neurotransmitter noradrenalin (the locus coeruleus-noradrenergic system). Dysregulation of this system contributes to cognitive dysfunction associated with a variety of psychiatric disorders, including attention-deficit hyperactivity disorder, sleep and arousal disorders, and post-traumatic stress disorder.[19] (**c**–**d**) Schematic representation of key cerebral circuits underlying cognitive function that are disrupted in psychiatric disorders (adapted from Millan et al.,[20] colors match the corresponding networks in Panel b). The brain structures showing the strongest clustering of perturbed genes are not only individually relevant to psychiatric disorders, as explained above, but they also constitute the core components of these integrated cognitive circuits, further supporting an etiological model where their dysregulation is a cause of impaired cognitive function in patients. (**c**) The frontal lobe, basal ganglia, and thalamus form a feedback loop that integrates cognitive control, attention, and working memory.[20] (**d**) Cognition is modulated by the cerebellum through feedback loops with the basal ganglia and the cortex, mainly via the thalamus and the pons.[20]

Nature Methods: doi:10.1038/nmeth.3799

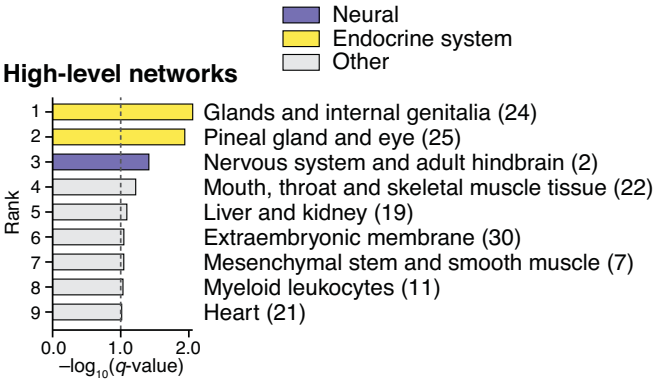
Supplementary Figure 34



Supplementary Figure 34: Connectivity enrichment for schizophrenia
Same as Fig. 3a, but showing all networks at FDR $< 10\%$.

Nature Methods: doi:10.1038/nmeth.3799

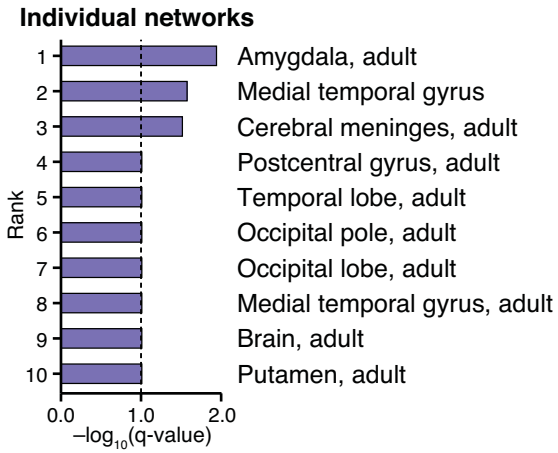
Supplementary Figure 35



Supplementary Figure 35: Connectivity enrichment for anorexia nervosa
Anorexia nervosa showed the strongest clustering of associated genes in high-level regulatory networks of the endocrine system (hormonal glands, Supplementary Fig. 14c) followed by nervous system and hindbrain, suggesting that endocrine dysregulation, a hallmark of chronic starvation in anorexia nervosa,[21] may be implicated.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 36

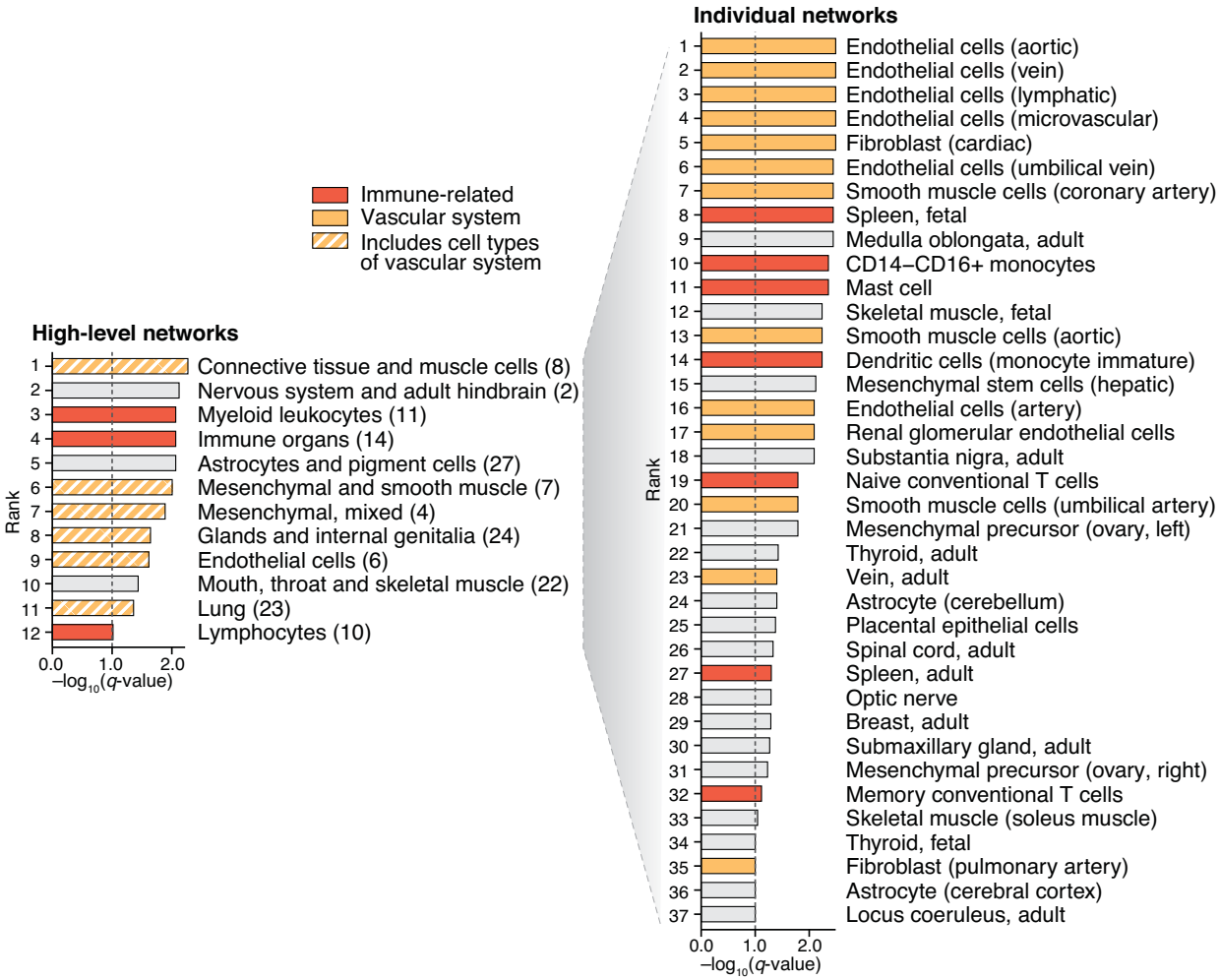


Supplementary Figure 36: Connectivity enrichment for bipolar disorder

Bipolar disorder did not show connectivity enrichment for the 32 high-level networks, but we nevertheless tested for enrichment in all individual networks of the nervous system (clusters 1–3 in **Supplementary Fig. 14**). All networks with score > 1.0 are shown (10% FDR, dashed line). The strongest clustering of perturbed genes is observed in the regulatory network of the amygdala, a group of nuclei playing key roles in memory modulation, emotional learning, emotional reactions, and mood. There are consistent findings in the literature suggesting that abnormalities in the structure and function of the amygdala are implicated in the etiology of bipolar disorder.[22]

Nature Methods: doi:10.1038/nmeth.3799

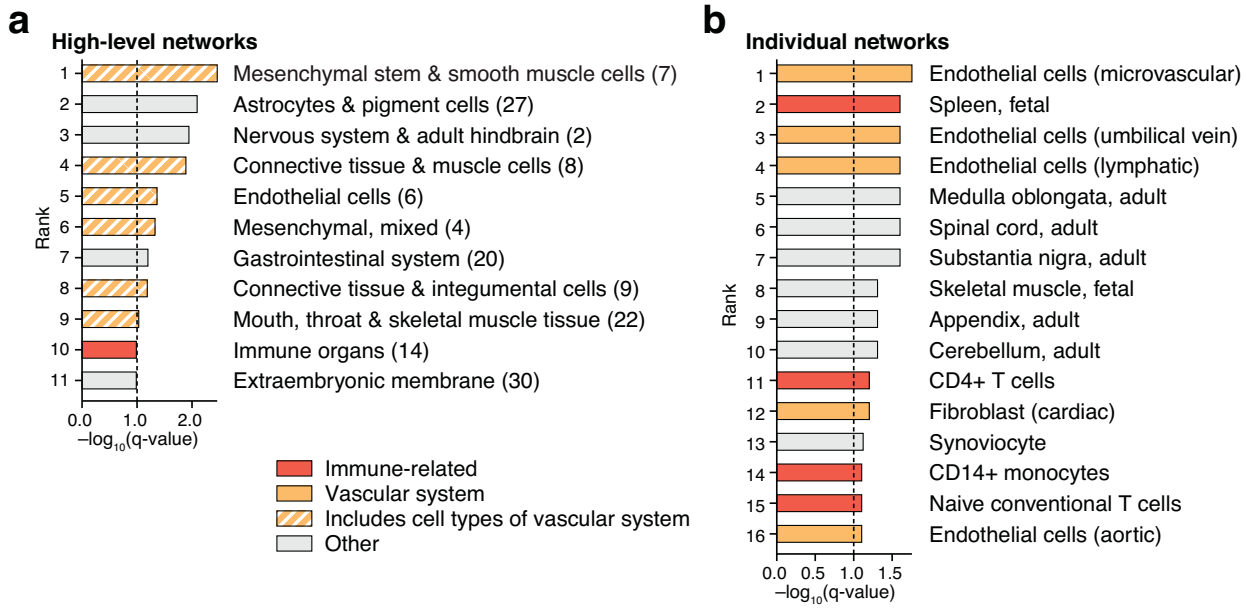
Supplementary Figure 37



Supplementary Figure 37: Connectivity enrichment for inflammatory bowel disease Same as Fig. 3b, but showing all networks at FDR < 10%.

Nature Methods: doi:10.1038/nmeth.3799

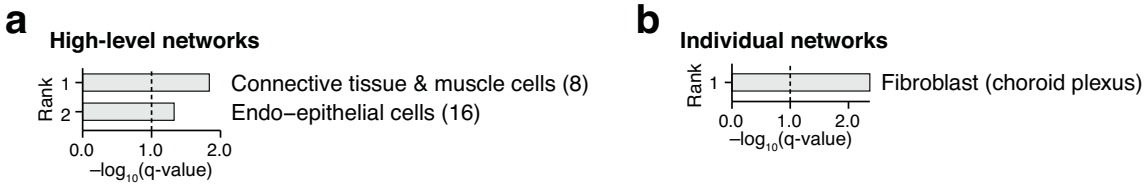
Supplementary Figure 38



Supplementary Figure 38: Connectivity enrichment for ulcerative colitis
Connectivity enrichment scores for ulcerative colitis across (**a**) the 32 high-level networks and (**b**) the corresponding individual networks. Results are highly consistent with those of inflammatory bowel disease, as expected (see Fig. 3b and main text). Namely, strong connectivity enrichment is observed for endothelial cells and immune organs. Endothelial cells drive pathophysiological processes including immune cell emigration and vascular growth (angiogenesis) that are direct or indirect targets of many drugs used to treat ulcerative colitis.[23] The spleen is the body's largest reservoir of monocytes and regulates inflammation through their storage and deployment.[24]

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 39



Supplementary Figure 39: Connectivity enrichment for Crohn's disease

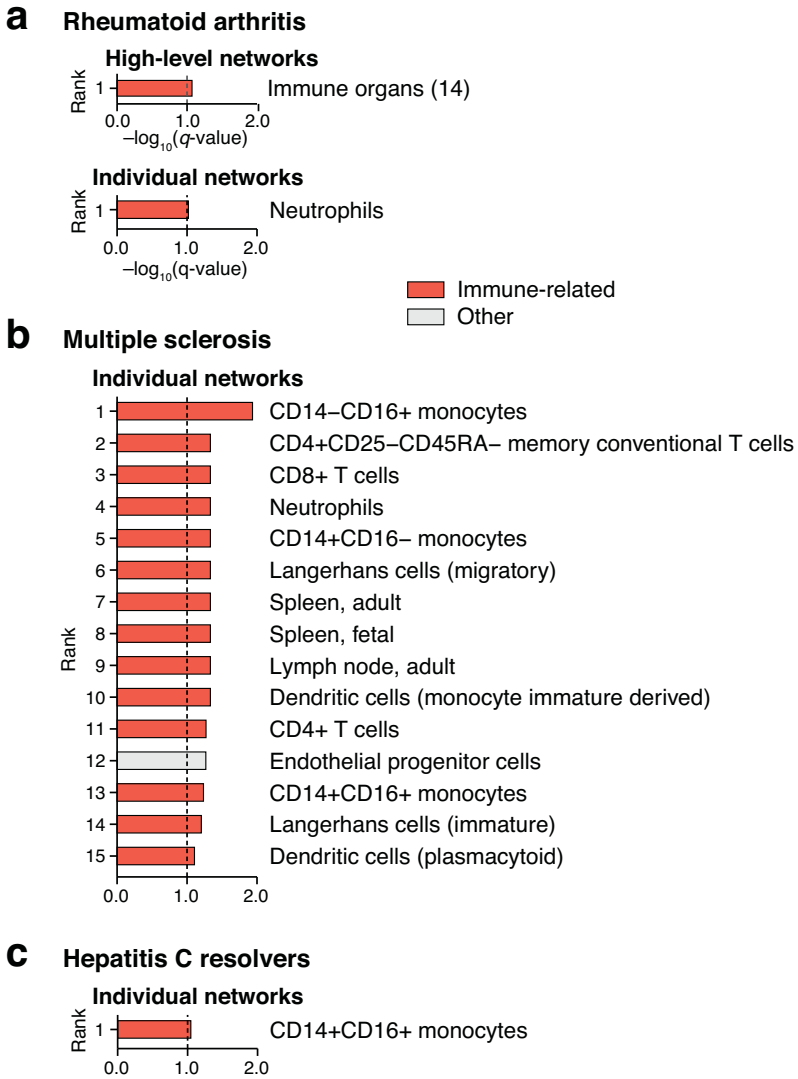
Connectivity enrichment scores for Crohn's disease across (**a**) the 32 high-level networks and (**b**) the corresponding individual networks. Connectivity enrichment is much weaker than for inflammatory bowel disease (IBD) and ulcerative colitis (UC) and does not seem to reveal relevant cell types.

(**a**) The high-level network for connective tissue & muscle cells (cluster 8) includes vascular cell types and also showed signal for IBD and UC. However, other clusters that include vascular cell types did not show signal here.

(**b**) The vascular cells showing strong clustering of perturbed genes for IBD and UC were not recovered here. Only one cell type shows significant connectivity enrichment (fibroblasts from the choroid plexus), which is likely a false positive.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 40



Supplementary Figure 40: Connectivity enrichment for rheumatoid arthritis, multiple sclerosis, and hepatitis C resolvers

Rheumatoid arthritis showed borderline connectivity enrichment in the high-level network of immune organs (**a**), while multiple sclerosis and hepatitis C resolvers did not show signal in any high-level network. We nevertheless tested these three immune-related traits for enrichment in all individual networks of the immune system.

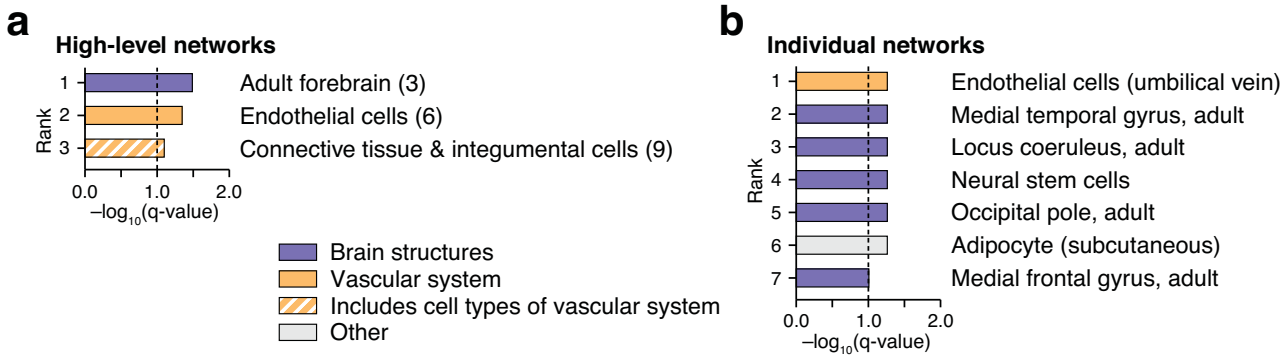
(**a**) For rheumatoid arthritis, the strongest evidence for perturbed regulatory modules was found in neutrophils, which have an activated phenotype in patients and contribute to pathogenesis through the release of cytotoxins, immunoregulatory molecules, and putative autoantigens.[25]

(**b**) Diverse immune cells showed connectivity enrichment for multiple sclerosis (MS), in particular T cells and monocytes. Indeed, the most consistent pathological finding in MS is the presence of T cell and monocyte/macrophage infiltrates among areas of myelin breakdown and gliosis in the central nervous system.[26]

(**c**) Hepatitis C resolvers, i.e., patients who spontaneously cleared the virus, showed borderline connectivity enrichment in monocytes, which have been proposed to spread infection over extended time because they get infected by the hepatitis C virus with few cytopathic effects.[27]

Nature Methods: doi:10.1038/nmeth.3799

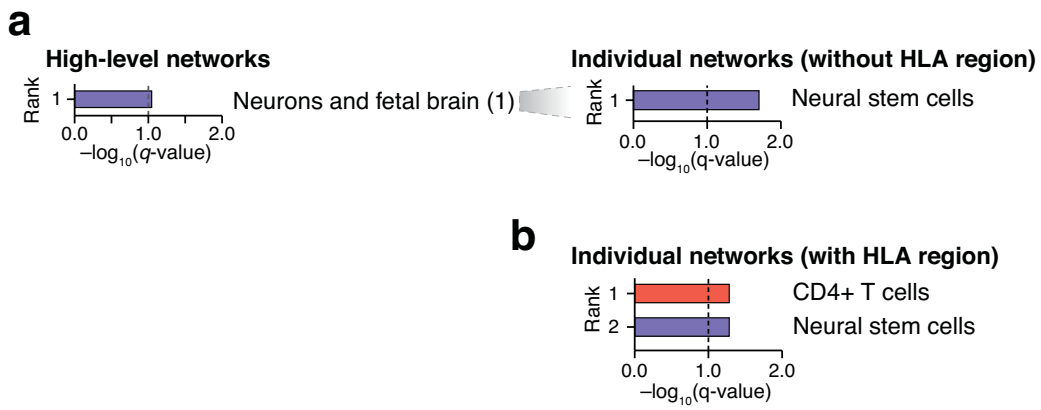
Supplementary Figure 41



Supplementary Figure 41: Connectivity enrichment for Alzheimer's disease
Connectivity enrichment scores for Alzheimer's disease across (**a**) the 32 high-level networks and (**b**) the corresponding individual networks.
(**a**) The strongest clustering of perturbed genes was found in regulatory networks of adult forebrain and endothelial cells, supporting the view that neurovascular dysregulation is implicated.[28] In contrast, these results do not confirm recent evidence for genetically driven dysregulation of immune cells in Alzheimer's disease,[29] however, the most relevant cell type (microglia) was not present in our library.
(**b**) Besides from endothelial cells mentioned above, several brain structures as well as neural stem cells showed connectivity enrichment. The medial temporal gyrus ranks second: medial temporal lobe atrophy is predictive of AD in subjects with mild cognitive impairment.[30] The locus coeruleus (3rd rank) is also particularly vulnerable in the early stages of AD and its degeneration contributes to disease pathology.[31]

Nature Methods: doi:10.1038/nmeth.3799

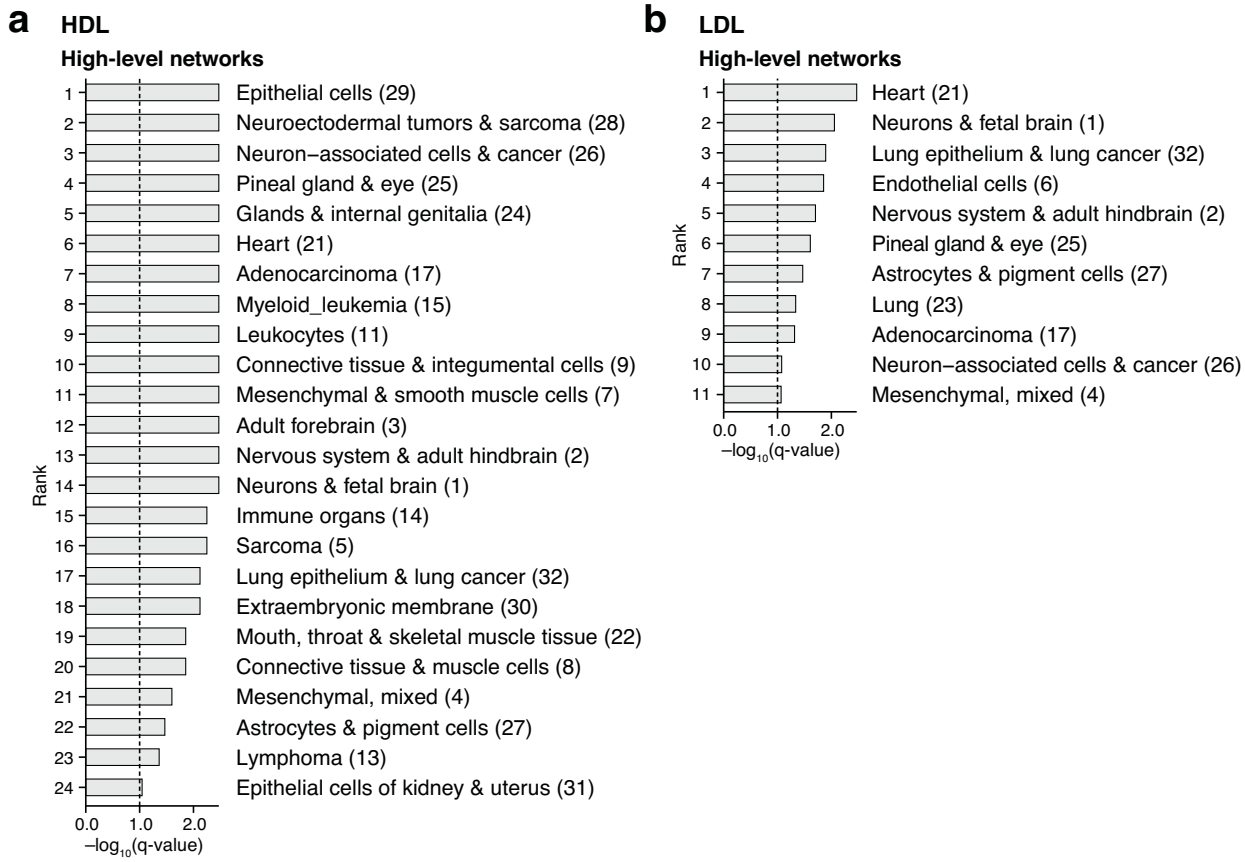
Supplementary Figure 42



Supplementary Figure 42: Connectivity enrichment for narcolepsy with and without HLA region Narcolepsy is a rare sleep disorder caused by an autoimmune attack targeting hypocretin-producing neurons.[32] We generally excluded the HLA region from the connectivity enrichment analysis because of its strong LD and association with many immune-related traits — this is a conservative measure often taken in network-based analyses of GWAS data to make sure that results are not driven by the HLA region[11] (see also Methods and **Supplementary Fig. 29**). (**a**) We observed connectivity enrichment only in regulatory networks of neural stem cells when the HLA region was excluded. This example shows that if the relevant cell type is not available (in this case the hypocretin neurons), networks of related cell types may still show signal.
(**b**) Given that the HLA region shows particularly strong association with narcolepsy,[32] we also run the connectivity enrichment analysis with the HLA region included (but still excluding all gene pairs that are in LD, see Methods) for all regulatory networks of the nervous and immune system. This analysis confirmed the observed clustering of perturbed genes in neural stem cells, while also showing signal for T cells, consistent with the autoimmune basis of narcolepsy and its association with the T-cell receptor alpha locus.[32,33]

Nature Methods: doi:10.1038/nmeth.3799

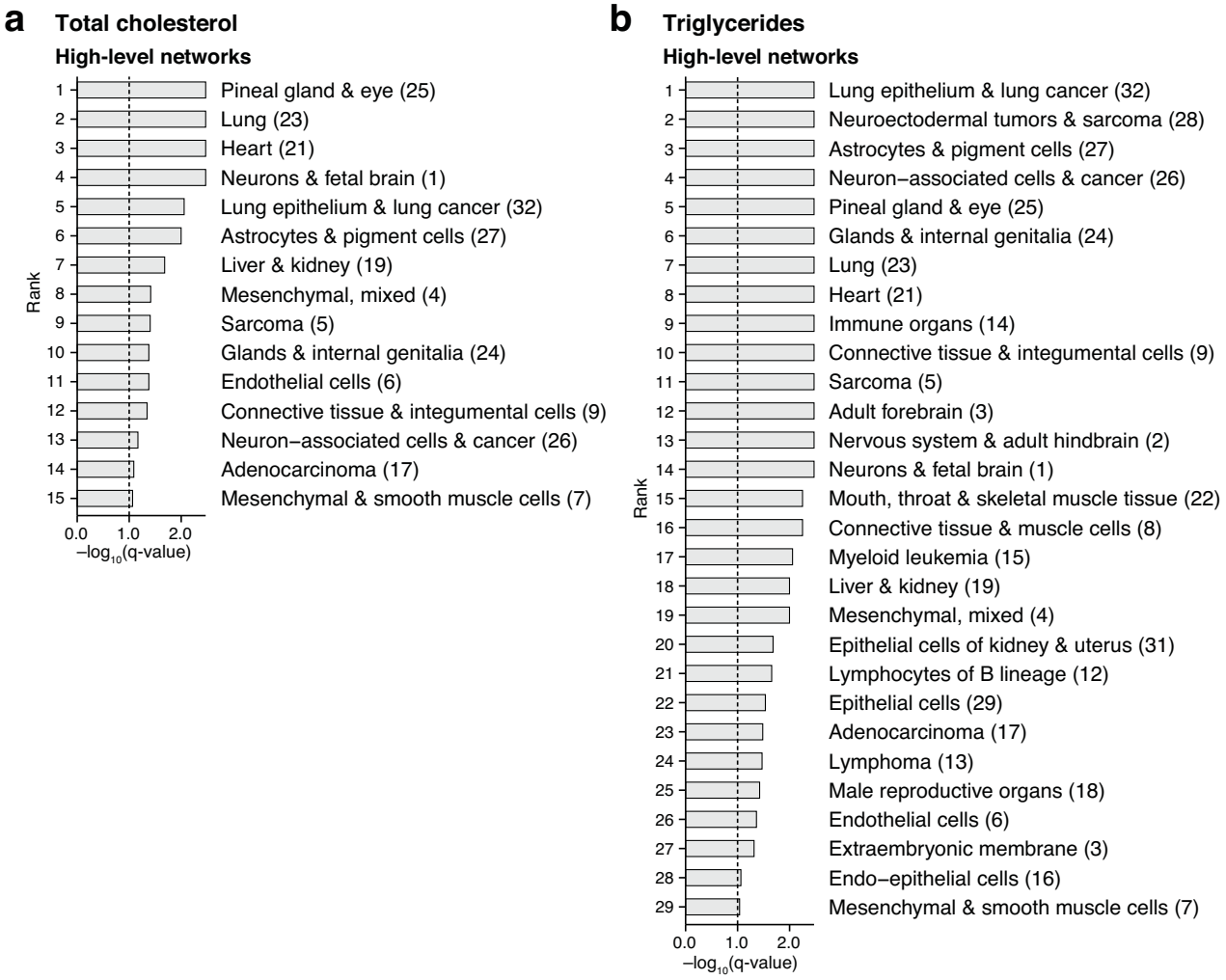
Supplementary Figure 43



Supplementary Figure 43: Connectivity enrichment for blood lipids (HDL and LDL)
Connectivity enrichment scores for high-density lipoprotein (HDL) and low-density lipoprotein (LDL) across the 32 high-level networks. We found strong clustering of perturbed genes in diverse high-level networks: over 75% of all networks show connectivity enrichment for HDL and over 30% for LDL. The same observations are made when testing individual networks: a large number of diverse cell types show connectivity enrichment (results not shown). Accordingly, the global regulatory network (obtained by taking the union of all cell type and tissue-specific networks, see Methods) also shows strong connectivity enrichment (**Supplementary Fig. 28**). Indeed, all human cells synthesize cholesterol, and thus share related pathways, because it is an essential component of cell membranes. Cells that regulate cholesterol levels in the blood, e.g., hepatocytes, likely rely on some of these shared cholesterol pathways, explaining the broad connectivity enrichment.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 44



Supplementary Figure 44: Connectivity enrichment for blood lipids (TC and TG)
Connectivity enrichment scores for total cholesterol (TC) and triglycerides (TG) across the 32 high-level networks. We found strong clustering of perturbed genes in diverse high-level networks: close to 50% of all networks show connectivity enrichment for TC and over 90% for TG. The same observations are made when testing individual networks: a large number of diverse cell types show connectivity enrichment (results not shown). This suggests that general pathways are implicated; see legend of **Supplementary Fig. 43** for details.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 45

c **Ontology enrichment (UBERON)**

Rank	Term	Q-value
1	embryo	0.0008
2	immune system	0.013
3	digestive system	0.017

Supplementary Figure 45: Connectivity enrichment for body mass index

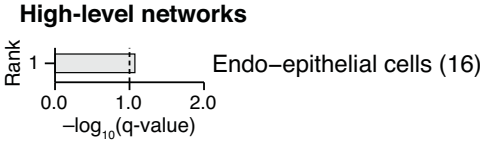
(Caption next page.)

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 45: (Previous page.)
Connectivity enrichment scores for body mass index (BMI) across (**a**) the 32 high-level networks and (**b**) the corresponding individual networks.
(**a**) BMI shows connectivity enrichment in diverse high-level networks, including those of the gastrointestinal system (rank 3) and immune organs (rank 4). Together with ulcerative colitis (**Supplementary Fig. 38**), BMI is the only trait showing connectivity enrichment for the gastrointestinal system.
(**b**) When testing individual networks, the strongest clustering of BMI-associated genes was found in networks of the digestive and immune system (red and purple indicate networks associated with the UBERON ontology terms *digestive system* and *immune system*, respectively, based on the sample annotation provided by FANTOM5). For the digestive system, we observed strong connectivity enrichment both for the lower gastrointestinal tract (small intestine, colon, rectum, gall bladder, appendix) and the upper gastrointestinal tract (esophagus, stomach) and related networks (tongue and tonsils, which group together with the esophagus in the hierarchical clustering of networks; **Supplementary Fig. 20**). For the immune system, we observed strong connectivity enrichment in different immune organs (tonsils, thymus, spleen, lymph nodes).
(**c**) The overrepresentation of digestive and immune system networks was confirmed through ontology enrichment analysis (besides the general term *embryo* (fetal samples), *immune system* and *digestive system* are the only significant terms; Methods).
These results suggest that perturbed pathways specific to the digestive system impact BMI. The digestive system is directly responsible for energy intake through (1) regulation of appetite by signaling to the brain and (2) digestion of nutrients. The strong signal for both the intestinal tract and the immune system may further suggest a potential link with the gut microbiome, which has recently been discovered to have a heritable component that impacts BMI.[34,35] Interestingly, previous studies by the GIANT consortium implicated mainly the central nervous system in obesity susceptibility (i.e., the receiver-end of appetite signals), and not the digestive and immune systems.[36] Reconciling these results is an important avenue for future work. Possibly, the regulatory circuit-based analysis presented here and pathway analyses employed by the GIANT consortium (e.g., DEPICT[37]) capture complementary aspects that may together give a fuller picture of tissues and biological pathways that impact BMI.

Nature Methods: doi:10.1038/nmeth.3799

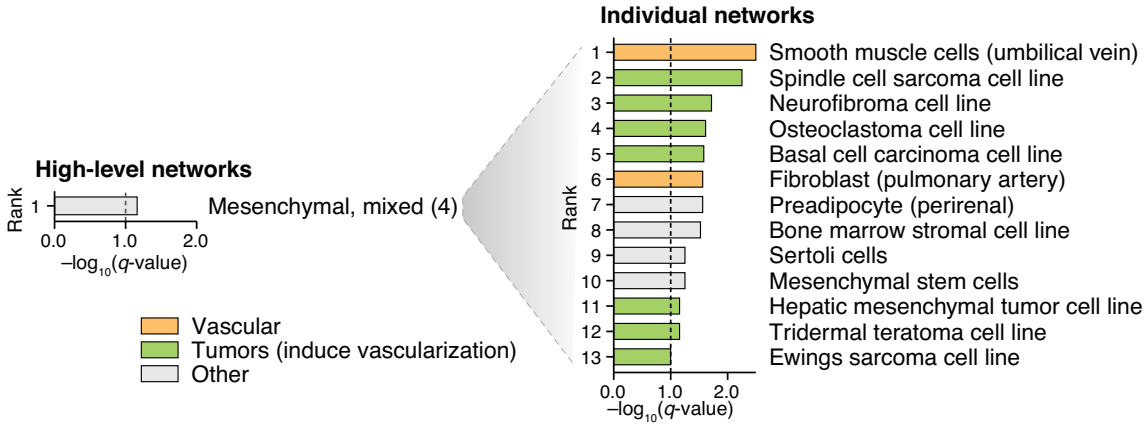
Supplementary Figure 46



Supplementary Figure 46: Connectivity enrichment for β-cell function
Connectivity enrichment scores for β-cell function (HOMA-β) across the 32 high-level networks. Only the network of cluster 16 (endo-epithelial cells) shows borderline enrichment. Given that none of the individual networks in this cluster is related to β-cells, our library does not include β-cells or pancreatic islets, and no other glycemic traits showed connectivity enrichment, this is likely a false positive.

Nature Methods: doi:10.1038/nmeth.3799

Supplementary Figure 47



Supplementary Figure 47: Connectivity enrichment for advanced macular degeneration (neovascular) Same as Fig. 3c, but showing all networks at FDR $< 10\%$. Age-related macular degeneration (AMD) of neovascular type shows the strongest connectivity enrichment in regulatory networks of vascular smooth muscle cells followed by diverse tumors, which induce vascularization to achieve growth. As a control, we further confirmed that the dry form of AMD, which does not involve neovascularization, does not show any connectivity enrichment in these networks.

Nature Methods: doi:10.1038/nmeth.3799

References

[1] Marstrand, T. T. & Storey, J. D. Identifying and mapping cell-type-specific chromatin programming of gene expression. *Proceedings of the National Academy of Sciences* **111**, E645–E654 (2014).

[2] Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).

[3] Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

[4] Roy, S. *et al.* A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Research* **43**, 8694–8712 (2015).

[5] Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* **14**, 283–291 (2004).

[6] Albert, R. Scale-free networks in cell biology. *Journal of Cell Science* **118**, 4947–4957 (2005).

[7] Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biology* **13**, R48 (2012).

[8] Lamparter, D., Marbach, D., Rico, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comp Bio* (in press).

[9] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

[10] Kondor, R. I. & Lafferty, J. D. Diffusion Kernels on Graphs and Other Discrete Input Spaces. In *Proceedings of the Nineteenth International Conference on Machine Learning*, ICML '02, 315–322 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002).

[11] Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genetics* **7**, e1001273 (2011).

[12] Erten, S., Bebek, G., Ewing, R. M. & Koyutürk, M. DADA: Degree-Aware Algorithms for Network-Based Disease Gene Prioritization. *BioData mining* **4**, 19 (2011).

[13] Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature biotechnology* **25**, 309–316 (2007).

[14] Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).

[15] Neph, S. *et al.* Circuitry and Dynamics of Human Transcription Factor Regulatory Networks. *Cell* **150**, 1274–1286 (2012).

[16] Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).

[17] Perez-Costas, E., Melendez-Ferro, M. & Roberts, R. C. Basal ganglia pathology in schizophrenia: dopamine connections and anomalies. *Journal of neurochemistry* **113**, 287–302 (2010).

[18] Hazlett, E. A. *et al.* Three-Dimensional Analysis With MRI and PET of the Size, Shape, and Function of the Thalamus in the Schizophrenia Spectrum. *American Journal of Psychiatry* (2014).

[19] Berridge, C. W. & Waterhouse, B. D. The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research. Brain Research Reviews* **42**, 33–84 (2003).

[20] Millan, M. J. *et al.* Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery* **11**, 141–168 (2012).

[21] Misra, M. & Klibanski, A. Endocrine consequences of anorexia nervosa. *The Lancet Diabetes & Endocrinology* **2**, 581–592 (2014).

[22] Garrett, A. & Chang, K. The role of the amygdala in bipolar disorder development. *Development and Psychopathology* **20**, 1285–1296 (2008).

[23] Cromer, W. E., Mathis, J. M., Granger, D. N., Chaitanya, G. V. & Alexander, J. S. Role of the endothelium in inflammatory bowel diseases. *World Journal of Gastroenterology* **17**, 578–593 (2011).

[24] Swirski, F. K. *et al.* Identification of Splenic Reservoir Monocytes and Their Deployment to Inflammatory Sites. *Science* **325**, 612–616 (2009).

[25] Wright, H. L., Moots, R. J. & Edwards, S. W. The multifactorial role of neutrophils in rheumatoid arthritis. *Nature Reviews Rheumatology* **10**, 593–601 (2014).

[26] Filion, L. G., Graziani-Bowering, G., Matusevicius, D. & Freedman, M. S. Monocyte-derived cytokines in multiple sclerosis. *Clinical and Experimental Immunology* **131**, 324–334 (2003).

Nature Methods: doi:10.1038/nmeth.3799

[27] Revie, D. & Salahuddin, S. Z. Role of macrophages and monocytes in hepatitis C virus infections. *World Journal of Gastroenterology : WJG* **20**, 2777–2784 (2014).

[28] Iadecola, C. Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nature Reviews. Neuroscience* **5**, 347–360 (2004).

[29] Gjoneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer/'s disease. *Nature* **518**, 365–369 (2015).

[30] Visser, P. J., Verhey, F. R. J., Hofman, P. a. M., Scheltens, P. & Jolles, J. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry* **72**, 491–497 (2002).

[31] Heneka, M. T. *et al.* Locus ceruleus controls Alzheimer's disease pathology by modulating microglial functions through norepinephrine. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 6058–6063 (2010).

[32] Hor, H. *et al.* Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy. *Nature Genetics* **42**, 786–789 (2010).

[33] Hallmayer, J. *et al.* Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nature Genetics* **41**, 708–711 (2009).

[34] Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).

[35] Sommer, F. & Bäckhed, F. The gut microbiota — masters of host development and physiology. *Nature Reviews Microbiology* **11**, 227–238 (2013).

[36] Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

[37] Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* **6**, 5890 (2015).

Nature Methods: doi:10.1038/nmeth.3799